Displacement or Complementarity? The Labor Market Impact of Generative AI*

Wilbur Xinyuan Chen
Hong Kong University of Science and Technology

Suraj Srinivasan
Harvard Business School
Digital, Data, Design Institute at Harvard

Saleh Zakerinia
The Ohio State University

November 2025

Abstract

Generative AI is poised to reshape the labor market, affecting cognitive and white-collar occupations in ways distinct from past technological revolutions. Exploiting the November 2022 release of ChatGPT and a near-universe of U.S. job postings linked to *O*NET* tasks, we examine how generative AI shifts hiring and skill requirements across occupations using synthetic difference-in-differences. After the release of ChatGPT, hiring falls in occupations with high labor-substitution potential (those with high automatable-task shares) by 13 percent and rises in those with high labor-augmentation potential (those with high automation-exposure diversity) by 20 percent. Skill requirements also move in parallel, declining in the former, while increasing in the latter occupations. Lastly, we analyze the unexposed tasks among occupations with high labor-augmentation potential, and we show that these tasks comprise of more social and technical skills, which highlights the potential human skills that complements Generative AI where automation-exposure diversity is greater.

Keywords: Generative Artificial Intelligence, Labor Market, Automation and Augmentation

JEL Codes: J23, J24, L86, O33

1 Introduction

Artificial intelligence (AI) has advanced rapidly in recent decades and is increasingly characterized as the next industrial revolution, with the potential to transform the economy and redefine the nature of work (Brynjolfsson and McAfee, 2017). The development and diffusion of generative AI technologies, particularly following the public release of ChatGPT in late 2022, has further accelerated this transformation. While scholars have argued that these technologies will have a profound impact on the labor market (Eloundou et al, 2024), the impact of this breakthrough on the labor market remains dynamic and complex. Unlike earlier technological revolutions, generative AI can perform sophisticated cognitive tasks such as problem-solving and decision-making that traditionally required human expertise. Thus, this novel capability of generative AI raises fundamental questions about which occupations face heightened substitution risks from these technologies, and which occupations stand to benefit from its complementary potential through enhanced worker productivity.

While many argue that generative AI, as a capital-augmenting technology, will displace workers by automating cognitive tasks previously performed by humans, thereby reducing labor demand (Eisfeldt et al., 2023; Frank et al., 2025), others suggest it could complement labor by boosting worker productivity and increasing labor demand (Agrawal et al., 2023b; Noy & Zhang, 2023). This could occur through productivity gains from automating parts of tasks within an occupation (Acemoglu & Restrepo, 2018a) as well as the creation of entirely new tasks enabled by generative AI (Acemoglu & Restrepo, 2019). The net effect of these opposing forces on labor demand determines how generative AI will ultimately impact the labor market, making the actual hiring outcomes in labor markets an empirical question.

Theoretical research on automation's labor market effects is well-established (Acemoglu & Restrepo, 2018a, 2019, 2020), but recent studies are only beginning to explore how generative AI may impact jobs, productivity, and firms (Eisfeldt et al., 2023; Eloundou et

al., 2024; Felten et al., 2023; Noy & Zhang, 2023). However, most of these studies rely on task-exposure models to assess automation (labor-substitution) potential, rather than analyzing realized changes in labor demand. Moreover, these studies have primarily focused on generative AI's potential to automate work and to substitute labor, and relatively little attention has been given to its potential to augment human labor by enhancing productivity in occupations with non-automatable tasks.

To address this gap, we construct new measures that capture the potential of generative AI to either automate or augment different occupations, based on their task structure. Using a near-universe dataset of U.S. job postings from *Lightcast*, we empirically validate these measures and analyze how generative AI is reshaping labor demand and skill requirements. While job postings do not measure realized employment, they provide granular, forward-looking signals of labor demand at the firm-occupation-quarter level and are widely used in economic research as proxies for firms' hiring behavior, especially for high-skill occupations (Hershbein & Kahn, 2018). With these job postings, we examine how firms adjust their demand for occupations with greater overall exposure to automation versus those more likely to benefit from augmentation, and how skill requirements evolve in response to generative AI adoption.

Our analysis begins by assessing the automation potential of the tasks associated with each occupation in the *Occupational Information Network (O*NET) v25.1* dataset. Following the methodology of Eloundou et al. (2024), we utilize OpenAI's ChatGPT (GPT-4o) to assess the potential for task automation by generative AI. Motivated by prior literature on the labor substitution and complementary effects of technological change (Acemoglu 2002b; 2003; Acemoglu & Autor, 2011), we construct two occupation-level measures based on the task-level automation potential of generative AI. The first measure captures the overall share of tasks likely to be automated (automatable-task share), and the second measure reflects the diversity of automatable and non-automatable tasks (automation-exposure diversity). Collectively, these

measures allow us to distinguish between occupations that are highly susceptible to substitution effects (i.e. those with high automatable-task share) and those that may benefit from productivity-enhancing augmentation (i.e. those with high automation-exposure diversity).

By identifying occupations with high automatable-task share and high automation-exposure diversity with *LightCast*, we study the changes in job postings before and after the introduction of ChatGPT in these occupational groups. Using Poisson difference-in-differences models weighted by synthetic control weights, we find that job postings for occupations in the top quartile of the automatable-task share declined by 13% per firm-occupation quartile per quarter, while postings for occupations in the top quartile of the automation-exposure diversity increased by 20%, relative to control occupations. This heterogeneous impact of generative AI on labor demand underscores its potential as a labor-augmenting force, contrasting prevailing concerns about its role in reducing overall labor demand[1]. This nuanced perspective contributes to the ongoing debate on AI's labor market impact, highlighting both the challenges and opportunities presented by this emerging technology.

Our next set of analyses focuses on the impact of generative AI on the skills associated with job postings. Utilizing the required skills data for each job posting from *LightCast*, we employ ChatGPT to classify skills and identify those exposed to generative AI. We hypothesize that the number of generative AI-exposed skills, total required skills, and newly required skills will decline in occupations with a high automatable-task share but will likely increase in those with a high automation-exposure diversity. This hypothesis aligns with the Acemoglu & Restrepo (2018b, 2019, 2020a) framework of task displacement versus task reinstatement, and

[1] For examples of prevailing concerns about AI-induced job displacement, see Goldman Sachs's estimate that generative AI could lead to the loss or degradation of up to 300 million jobs in the U.S. and Europe (Forbes, March 31, 2023), and Geoffrey Hinton's warning that AI may cause "massive unemployment" while boosting corporate profits (Fortune, September 6, 2025).

with Autor et al. (2024)'s emphasis on AI-driven complementarity, where automation reduces skill demands in some occupations while augmentation expands them in others.

Our findings support for this hypothesis. Following the introduction of ChatGPT, we observe a 7% quarterly decrease in generative AI-exposed skills per firm-occupation quartile among jobs in the top quartile of automatable-task shares. In contrast, there is an 8% quarterly increase in generative AI-exposed skills per firm-occupation quartile for jobs most susceptible to augmentation (in the top quartile of the automation-exposure diversity). Similar trends are found in both the total number of required skills and newly required skills. Overall, these results underscore the heterogenous impact of generative AI on skill requirements across occupations. They suggest that while generative AI reduces the need for certain skills in highly automatable occupations, it simultaneously creates demand for new skills in occupations that can be effectively augmented by generative AI technologies.

We also find that the labor market effects of generative AI vary substantially across firms and industries. Job growth associated with high automation-exposure diversity occupations is concentrated in large and public firms, while declines in high automatable-task occupations are strongest in tech-intensive sectors such as finance and technology. In contrast, small firms and government organizations show weaker or no measurable response. These patterns suggest that the benefits and disruptions of generative AI are not evenly distributed and are perhaps shaped by organizational absorptive capacity and sector-specific technological adoption dynamics (Cao & Iansiti, 2022).

Lastly, we analyze the characteristics of tasks that are unexposed in high automation-exposure diversity occupations to understand the relevant human skills that could complement generative AI technologies. To that end, we estimate linear probability regression of an indicator of unexposed tasks among occupations on the skills characteristics of these tasks. These analyses demonstrate that unexposed tasks, particularly among high automation-

exposure diversity occupations, comprise of more social and technical skills compared to basic skills, and highlight that these skills may serve as complements to generative AI.

Our approach builds on canonical models of technological change, such as Autor et al. (2003) and Acemoglu & Restrepo (2019), which emphasize that the composition of tasks within occupations determines whether technology substitutes for or complements human labor. We extend this framework to generative AI by distinguishing between occupations with a high automatable-task share versus those with a high automation-exposure diversity, based on their task content. Our task-level classification thus enables us to empirically test how generative AI affects labor demand along theoretically grounded dimensions. In doing so, our analysis also speaks to the literature on skill-biased technical change (Katz & Murphy, 1992), highlighting how generative AI may shift labor demand toward more complex, skill-intensive tasks where human-AI complementarity is greatest.

This study also contributes to multiple strands of the academic literature. First, it builds on research examining how technological change affects the labor market by empirically examining whether generative AI leads to labor displacement or augmentation. Technological advancements can be factor-biased, augmenting either capital or labor (Acemoglu et al., 2024; Acemoglu & Restrepo, 2019), and may also create new tasks that reshape workforce composition and skill requirements (Acemoglu & Restrepo, 2019; Autor et al., 2022). Historically, automation has favored capital and reduced labor's share of income, contributing to wage stagnation, particularly in manufacturing (Acemoglu & Restrepo, 2019). However, due to its distinct capabilities, some researchers argue that AI may diverge from past technologies by increasing demand for occupations with low automatable-task shares (Fossen et al., 2022). Others suggest that AI could replicate traditional automation's displacement effects, reducing overall labor demand (Acemoglu et al., 2022; Agrawal et al., 2023a). This study contributes to this debate by providing the novel empirical evidence on the displacement

and augmentation effects of generative AI, using a near-universe dataset of U.S. job postings and task-level automation measures.

Second, this study contributes to the nascent and rapidly growing literature on the economic consequences impact of AI. Notably, research on AI's labor market impact is expanding rapidly (e.g., Acemoglu et al., 2022; Agrawal et al., 2023a; Fossen et al., 2022; Frank et al., 2025; Green, 2024).[2] However, many of these studies provide limited empirical insight into AI's heterogeneous impact on labor market and specifically, AI's potential to augment rather than replace human labor in certain occupations. By developing a measure for automation-exposure diversity and distinguishing it from the automatable-task share, our study offers a more granular and complete view of generative AI's impact on labor market effects.

Third, our research contributes to recent research on the productivity impact of generative AI in specific occupations. Many of these studies focus on controlled or specialized settings, such as customer support roles (Brynjolfsson et al., 2023), the accounting profession (Choi & Xie, 2025) or online platforms (Hui et al., 2024; Liu et al., 2023). Others employ lab-based experiments (e.g., Doshi & Hauser, 2024; Noy & Zhang, 2023), to more precisely evaluate the productivity impact in certain occupations. While these studies find productivity gains in specific settings, they stop short of offering a comprehensive, economy-wide view of generative AI's impact on labor demand and skills. We address this gap by analyzing the labor-substitution and -augmentation impact of generative AI across a wide range of occupations. Building on the task-based index developed by Eloundou et al. (2024), which has been used to study generative AI's economic impact (Acemoglu, 2024; Berger et al., 2024; Eisfeldt et al., 2023), we introduce an automation-exposure diversity index that capture generative AI's labor-

[2] Much of this literature relies on automation indices developed by Felten et al. (2018, 2023), Brynjolfsson & Mitchell (2017), Brynjolfsson et al. (2018), Webb (2019), and Eloundou et al. (2024), to measure AI's impact at the firm or industry level, focusing on outcomes such as employment, wage inequality, and productivity.

augmenting potential alongside its labor-substitution risks. By applying these indices to a near-universe dataset of U.S. job postings, we provide empirical insights into both the displacement and augmentation effects of generative AI across all occupations in the U.S. economy.

2 Conceptual Framework

2.1 Labor Market Implications of Generative AI

The literature widely acknowledges that technological change can be either capital-intensive, replacing labor through task automation, or labor-intensive, increasing labor demand by creating new tasks and boosting productivity (Acemoglu, 2002a; Acemoglu & Restrepo, 2018b). Generative AI differs from earlier waves of technological change by primarily targeting non-routine cognitive tasks in white-collar jobs, rather than routine cognitive or manual blue-collar work. This dual capacity—automating non-routine cognitive tasks while complementing human capabilities—creates heterogeneous effects across occupations, contingent on their task composition.

Occupations predominantly composed of automatable tasks are likely to experience job displacement as generative AI substitutes human labor. This aligns with the traditional framework of automation, where technological advancements reduce labor demand by replacing routine and codifiable tasks (Acemoglu & Restrepo, 2019; Autor et al., 2003). Generative AI's proficiency in tasks such as text generation, summarization, language translation, and even non-codifiable tasks diminish the need for human input in a potentially new and wider set of high labor-substitution potential roles, leading to a decline in job postings in a broad set of occupations.

Conversely, occupations that mix automatable and non-automatable tasks are more likely to benefit from generative AI through productivity enhancements. In these occupations, generative AI can augment human labor by improving task efficiency, allowing workers to focus on higher-value activities that require human judgment, creativity, and problem-solving

(Acemoglu & Autor, 2011). As productivity increases among occupations that are conducive to augmentation, firms may expand hiring to capitalize on these productivity gains, potentially leading to an increase in job postings. At the same time, if existing workers can take on more responsibilities due to freed-up time from automated tasks, firms may not need to expand headcount. While both channels are theoretically possible, prior work suggests that augmentation effects tend to dominate in occupations with diverse task structures, as technology often frees workers from routine elements and shifts demand toward more complex, non-routine tasks (Autor et al., 2003; Acemoglu & Autor, 2011; Acemoglu & Restrepo, 2019). This motivates our first hypothesis.

H1: Generative AI reduces job postings in occupations with high automatable-task share (i.e. high labor-substitution potential) and increases job postings in occupations with high automation-exposure diversity (i.e. high labor-augmentation potential).

2.2 Impact of Generative AI on Skills

Technological change has historically exhibited a skill-biased focus, favoring skilled workers while diminishing demand for routine tasks (Autor et al., 2003, 2006). Previous waves of automation primarily displaced blue-collar occupations, where machines substituted for routine physical tasks, while later waves displaced routine cognitive roles such as administrative work. In contrast, generative AI marks a qualitative departure by targeting *non-routine cognitive* tasks that have historically been the domain of highly educated, white-collar workers, such as language translation, text generation, coding, and problem-solving: areas previously considered less susceptible to automation (Ellingrud et al., 2023).

At the same time, generative AI also has the potential to augment labor in occupations that require creativity, judgment, and advanced reasoning. Thus, its impact on skills demand is not uniform but depends critically on the task composition of occupations.

2.2.1 Effect on Occupational Skill Requirements

Generative AI reshapes skills demand by both reducing the set of required capabilities, while expanding others, depending on the task composition of occupations. In occupations with a high automatable-task share, generative AI can directly substitute for a wide range of cognitive tasks, including not only routine codifiable ones but also more complex tasks such as drafting, summarization, or translation. This substitution simplifies workflows and reduces the average number of required skills, including generative AI-exposed skills. This perspective reflects theories of deskilling, in which automation narrows the scope of job responsibilities and consolidates competencies (Autor et al., 2003; Acemoglu & Restrepo, 2019).

In contrast, in occupations with high automation-exposure diversity, generative AI likely complements workers by augmenting their productivity. This leads to an increase in both the average number of required skills and generative AI-exposed skills, reflecting a process of upskilling (Acemoglu & Restrepo, 2019; Bessen, 2018). For example, a physician interpreting diagnostic tests may rely on generative AI to handle more standardized components of image analysis, but the physician's expertise remains essential for contextualizing findings, confirming diagnoses, and determining treatment plans. In such cases, workers must integrate generative AI outputs into broader workflows, which increases the salience of complementary capabilities such as tool proficiency, judgment, and creative problem-solving. Rather than reducing demand for these skills, generative AI amplifies their importance and broadens overall skill requirements (Noy & Zhang, 2023).

H2: Generative AI simplifies skill requirements in occupations with a high automatable-task share but expands them in occupations with high automation-exposure diversity, increasing both the average number of required skills and generative AI-exposed skills.

2.2.2 Effect on Newly Required Skills

As argued by Autor et al. (2024), new work emerges in response to technological shocks that alter task structures and create opportunities for task innovation. Generative AI represents such a shock by introducing new capabilities into existing workflows that are predicted to impact occupations in different ways.

In occupations with high automatable-task shares, the scope for new skill requirements diminishes. As more cognitive tasks are automated, job functions become standardized and consolidated, leaving fewer opportunities for experimentation and skill development.

By contrast, in occupations with high automation-exposure diversity, generative AI complements existing work by enabling new workflows and augmenting complex tasks, which spurs demand for new skills. These skills are often tied to the effective integration of generative AI into work processes. For example, tool proficiency, creative reasoning, and higher-order problem-solving. As firms adopt generative AI, workers in these occupations are expected to continuously expand their capabilities, driving a process of adaptive upskilling.

H3: Generative AI dampens (accelerates) the emergence of new required skills in occupations with high automatable-task shares (high automation-exposure diversity).

3 Data and Measurement

3.1 Data Sources

We use two primary data sources in our study: task descriptions from *the Occupational Information Network (O*NET) v 25.1* database and U.S. job postings from *LightCast*, covering the period from 2019 to March 2025.

The *O*NET* dataset maintained by the U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). This dataset uses the O*NET-SOC taxonomy, which is based on the Standard Occupational Classification (SOC) system to classify different occupations. The dataset contains 19,265 tasks, and their associated task descriptions, for 923

occupations of U.S. workers, and also provides a rich set of variables for each occupation that describes the nature of work and worker characteristics.

The *LightCast* dataset captures the near-universe set of U.S. online job vacancies from roughly 40,000 company and recruiting websites. It includes detailed information on job titles, and required skills, enabling us to track changes in labor and skills demand over time.

3.2 Automatable Task Share and Automation Exposure Diversity Indices

To create our exposure indices, we use task descriptions from the *O*NET* dataset. Following the methodology established by Eloundou et al. (2024), we calculate each occupation's automatable task-share and automation-exposure diversity by first using OpenAI's GPT-4o model to assess whether each task can be effectively performed by generative AI.

In our prompt, we ask ChatGPT to assign each task to one of the following categories based on its potential exposure to large language models (LLMs):

E0: No exposure to LLMs.
E1: Direct exposure to LLMs, where access to an LLM can reduce task completion time by at least half without compromising quality.
E2: Exposure via LLM-powered applications, where the task cannot be accelerated by direct access to an LLM alone but can be completed more efficiently if supported by additional software developed on top of the LLM.
E3: Exposure given image capabilities, where completion time can be reduced by at least half if the LLM is combined with E2 software and systems capable of reading, creating, and interpreting images.

This classification rubric (detailed in Appendix A) is designed to capture the extent to which generative AI is able to perform each task. Following Eloundou et al. (2024), we compute an occupation-level automatable-task share by assigning numerical weights of 1 for E1, 0.5 for E2, and 0 for E0 and E3 for each task. These values are then weighted by each task's importance score in *O*NET* to emphasize tasks central to an occupation. In the Appendix, we also report results using alternative weighting schemes, such as treating E2 or E3 differently.

Importantly, while the classification rubric measures whether a task can be performed by generative AI, it does not distinguish between its implication for substitution and augmentation of labor at the occupation-level. For instance, tasks labeled E1 may be fully automatable or may still require human oversight—i.e., they may lead to productivity enhancement (augmentation) rather than displacement (automation) for different occupations. Thus, exposure at the task-level inherently blends both automation and augmentation potential.

To separate the potential labor substitution effects from labor-augmentation effects at the occupation level, we construct two occupation-level indices:

Automatable-task share: The importance-weighted mean exposure score across all tasks in an occupation. A higher score indicates that a larger share of tasks is directly exposed to generative AI (via E1 or E2), implying greater substitution potential.

$$\text{Automatable-Task Share Index}_j = \frac{\sum_{i \in j} \text{Importance}_i \cdot \text{Exposure Score}_i}{\sum_{i \in j} \text{Importance}_i}$$

Automation-exposure diversity: a normalized Herfindahl-Hirschman Index (HHI) capturing the dispersion of exposure types within an occupation—specifically, between AI-exposed (E1-E3) and non-exposed (E0) tasks. A higher index implies greater heterogeneity in exposure, consistent with the idea that some tasks may be automated while others require human input. This heterogeneity creates scope for augmentation.

$$\text{Automation Exposure-Diversity Index}_j = 1 - \left(\text{Share exposed}_j^2 + \text{Share non-exposed}_j^2\right)$$

This occupation-level separation allows us to disentangle labor-substitution from labor-augmentation effects (even though the underlying task-level measure tracks generative AI's overall automation potential). Occupations with high automatable-task shares are more likely to face labor-substitution pressures, while those with high automation exposure-diversity are more likely to benefit from generative AI as a labor-augmenting tool.

3.3 Skills Classification

In addition to occupation-level classification, we also classify skill requirements to study the impact of generative AI on new and generative AI-related skills. To measure generative AI-related skills, we employ ChatGPT to classify such skills in *Lightcast*. While traditional methods rely on keyword-based methods to measure technology exposure (e.g., Acemoglu et al., 2022 and Alekseeva et al., 2021), these methods are less suitable for measuring generative AI exposure as these technologies affects a broad spectrum of skills, including those indirectly related to AI, such as writing. *LightCast* also contains over 33,000 unique skills, making comprehensive and accurate classification essential. To address this challenge, we leverage ChatGPT's advanced language understanding capabilities to categorize skills into four distinct classes, as shown below. The full rubric for this classification is provided in Appendix A.

S0: Skills irrelevant to generative AI, with limited impact.
S1: Generative AI-relevant skills, essential for developing AI technologies.
S2: Skills complemented by generative AI, fundamentally human tasks that can be enhanced by AI.
S3: Skills substituted by generative AI, tasks that can be fully performed by AI.

For empirical analysis, we aggregate S1, S2, and S3 into a single category of generative AI-exposed skills, while S0 skills are classified as non-exposed. This aggregation facilitates a clean distinction between skills likely affected by generative AI and those that are not.[3]

Next, we identify new skills requirements at the firm-occupation level by defining a baseline set of required skills posted for each firm-occupation between 2015 and 2018. Starting from 2019, we track required skills on a quarterly basis; any skill appearing within a firm-occupation that was not present in the baseline set is classified as a new skill. This new skill is

[3] Table S1 shows that approximately 49% of skills in the dataset are not exposed to generative AI, while 51% have some level of AI exposure. This distribution highlights the broad potential reach of generative AI across a variety of skill types, underscoring its significant influence on workforce skill requirements.

then added to the baseline set for tracking in subsequent quarters. We repeat this process until the first quarter of 2025 to capture the emergence of new skills for each firm-occupation.

Lastly, in the main analysis, the skill measures are aggregated by first counting the number of distinct skills listed in each job posting, then averaging these values across all postings within each firm-occupation quartile-quarter. This approach captures the number of skills per posting, while avoiding confounding from variation in the number of job postings.

4 Research Design

4.1 Sample Construction

Using this data, we aggregate job postings and required skills at the firm-occupation quartile-quarter level, beginning in 2019. To capture treatment variation, we divide occupations into quartiles based on the distribution of their automatable-task share or automation-exposure diversity scores. For each firm, we then compute the total number of job postings and skills by quartile and quarter. This aggregation enables us to analyze trends in job and skills demand, based on occupation-level exposure scores, while preserving the treatment definition, reducing the dimensionality of the data, and facilitating firm-level interpretation.[4]

To improve the reliability of the sample, we apply several consistency filters. First, we restrict the analysis to firms that post at least 10 job postings per quarter across all occupations, ensuring a minimum level of posting intensity. Second, we require each firm-occupation-quartile to appear in at least one quarter of every year between 2019 and 2025, so that each

[4] For robustness, we also estimate models at the occupation-quarter level and at the firm-occupation-quarter level for a subset of firms, where smaller sample size allows for more granular estimation.

firm-occupation-quartile is observed in each of the years.[5] Collectively, these restrictions yield a sample of firms with sustained and non-sparse posting activity, which reduces noise from sporadically active employers and yielding balanced analysis of employment trajectories.

4.2 Identification Strategy: Launch of ChatGPT

We exploit the public release of ChatGPT in November 2022 as a plausibly exogenous shock to the availability of generative AI in the workplace. This event marked the first widely-accessible deployment of large language models (LLMs), providing a quasi-natural experiment to evaluate the labor market impact of generative AI technologies. The sudden release and the immediate surge in public attention make it a suitable inflection point for identifying shifts in labor demand and skill requirements. As shown in Figure 1, search interest in generative AI and LLMs spiked sharply after ChatGPT's release, with Google Trends indices rising dramatically starting in the fourth quarter of 2022. This discontinuous jump in public awareness supports our interpretation of ChatGPT's launch as a salient and plausibly exogenous shock, which underpins our empirical strategy grounded in occupational exposure to generative AI.

Our empirical identification strategy is based on a difference-in-differences (DiD) design, by comparing outcomes before and after the launch of ChatGPT across occupations with differing levels of pre-treatment exposure to generative AI. Treatment is defined based on exposure quartiles constructed separately for automatable-task share and automation-exposure

[5] We begin the sample in 2019 to ensure sufficient pre-treatment history while minimizing the confounding influence of early-pandemic volatility. Although the study period includes COVID-19 shock, the timing of the treatment event and the design of the empirical strategy, described in the next section, are intended to distinguish the effects of generative AI from broader macroeconomic or pandemic-related shifts.

diversity indices (as described in Section 3).[6] For each index, the treated group comprises occupations in the top quartile of the index, while the control group consists of occupations in the third and fourth quartiles. The second quartile is excluded to avoid contamination from borderline exposure cases.

To better identify the DiD treatment effects of the ChatGPT shock, we employ the synthetic difference-in-differences (SDID) estimator developed by Arkhangelsky et al. (2021). SDID combines features of synthetic control and traditional DiD to construct a weighted control group that closely matches the pre-treatment trajectory of the treated group, thereby relaxing the parallel trends assumption. While treatment is defined at the occupation level, occupations differ not only in their exposure to generative AI but also in their frequency across firms and over time. Such group size imbalance is a concern for standard DiD models, which apply equal weights across units and may therefore overrepresent smaller or less comparable control units.[7] SDID mitigates this issue by optimally re-weighting control units and time periods to better match the treated group's pre-treatment trajectory, yielding more credible

[6] One potential concern with this treatment assignment is reverse causality, that is, occupations with rising or declining demand coincidentally appear more "exposed" to generative AI. We mitigate this concern in several ways. First, our treatment assignment is based on exposure scores derived from a standardized task-level rubric applied to *O*NET* data, which predate the public release of ChatGPT, and are fixed across time and independent of labor market outcomes. While we use a post-2022 large language model (GPT-4o) to generate exposure labels, we ask it to evaluate the technical characteristics of tasks, not occupational demand or employment trends. Second, the treatment timing is externally determined: ChatGPT's release in November 2022 represents an exogenous, and widely publicized event, unlikely to be anticipated by most firms or workers in advance (Figure 1).

[7] For example, the summary statistics reported in Table 2 show that the average number of postings for top-quartile occupations is comparable to, or even exceeds, the overall sample mean, suggesting compositional differences in hiring patterns of treatment and control groups.

estimates.[89] In doing so, this procedure also mitigates concerns that the occupations in the top quartile of exposure differ systematically from those in the lower quartiles.

In our main analysis, we construct a balanced panel of *N* firm-occupation-quartile units observed in every quarter between 2019 and 2025. Each unit *i* corresponds to a firm *f* and an occupation quartile *q* (based on occupation-level sorts on the automatable-task share or automation-exposure diversity indices), and the outcome $\boldsymbol{Y_{it}}$ is the number of job postings (or associated skill measures) in quarter *t*. For units with no postings in a given quarter, we record a zero value, ensuring that all eligible units are represented in every period and that the estimation sample takes the form of a balanced panel. The SDID procedure estimates parameters by solving the following optimization problem:

$$\left(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \boldsymbol{arg\ min}_{\tau,\mu,\alpha,\beta}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \mu - \alpha_i - \beta_t - \boldsymbol{Post\ Treatment}_{it}\tau)^2\ \widehat{\omega_i}\widehat{\lambda_t}\right\}$$

[8] The SDID estimator is particularly useful in this context because it addresses the key identification concern of non-parallel pre-treatment trends. Unlike traditional DiD, which assumes treated and control groups would have followed similar trajectories in the absence of treatment, SDID relaxes this assumption by constructing unit and time weights to align them in the pre-period. This extension of the synthetic control method (Abadie et al., 2010) to settings with many treated units and time periods produces more reliable estimates in large panels with potential heterogeneity in pre-treatment dynamics.

[9] In additional robustness analysis, we complement the SDID analysis by estimating conventional difference-in-differences models with propensity score matching. Specifically, we match treated and control occupations based on a set of pre-treatment covariates drawn from *O*NET*, including characteristics such as required education, adaptability, stress tolerance, innovation, and task dependency, among others. These features capture heterogeneity in occupational complexity, cognitive demands, and work context. We then estimate DiD models at the firm-occupation level using the matched sample. In additional robustness analysis, we also study a restricted sample of firms that employ both treated and control occupations, allowing us to identify treatment effects based solely on within-firm variation.

where, τ is the average causal effect of the treatment, μ is an intercept, $\boldsymbol{\alpha_i}$ and $\boldsymbol{\beta_t}$ are unit and time fixed effects, and $\boldsymbol{Post\ Treatment_{it}}$ indicates whether unit *i* is treated in period *t*. Unit weights $\widehat{\boldsymbol{\omega_i}}$ are optimally selected with the SDID procedure such that the weighted control group closely matches the treated group's average pre-treatment trajectory. Time weights $\widehat{\boldsymbol{\lambda_t}}$ are chosen through the procedure to align pre- and post-treatment periods, thereby preserving consistent differences for the control group. Thus, by constructing a synthetic control group tailored to the treated units, this method relaxes the strict parallel-trends assumption required in standard DiD.[10][11]

As an alternative specification, we also apply a Poisson DiD model, which has a functional form that is better suited for count data such as job postings (Cohn et al., 2022). To account for differences in pre-treatment trends, we weight control units using synthetic weights obtained from SDID, ensuring that pre-treatment trends between treated and control groups are balanced. The model is specified as follows:

[10] Under general conditions, including cases where treatment assignment correlates with unit-level time trends and treatment effects vary across units, Arkhangelsky et al. (2021) show that SDID yields a consistent and asymptotically normal estimate of the average treatment effect, provided there are sufficient control units and pre-treatment periods relative to treated units and post-treatment periods. In our study, these likely assumptions hold, as each firm's treated quartile (quartile 1 of the automatable-task share for the automation analysis, or quartile 1 of the automation-exposure diversity for the augmentation analysis) has two control quartiles (quartiles 3 and 4), and the pre-treatment period spans from 2019 to the third quarter of 2022, and the post-treatment period extends from the fourth quarter of 2022 to the first quarter of 2025.

[11] To examine how treatment effects evolve over time and to verify pre-treatment parallel trends, we construct an event study-style plot using SDID, following the approach in Clarke et al. (2023). This approach plots the difference in the average outcome between treatment and synthetic control units, relative to the optimally weighted pre-treatment means. To construct confidence intervals, we use block-bootstrap resampled standard errors. For more details see Arkhangelsky et al. (2021).

$$\mathbf{E}(\boldsymbol{Y_{it}}|.) = \mathbf{exp}\,(\boldsymbol{\tau Post\ Treatment_{it}} \times \boldsymbol{Treated_i} + \boldsymbol{\alpha_i} + \boldsymbol{\beta_t} + \boldsymbol{\epsilon_{it}})$$

where $\boldsymbol{Treated_i}$ is an indicator variable identifying whether group *i* belongs to the treated group (quartile 1 of automatable-task share or automation-exposure diversity, depending on the analysis). In this specification, τ approximates the percentage change in the outcome variable for the treated group post-treatment, compared to the control group post-treatment. By weighting control units using SDID-derived synthetic weights, this approach mitigates biases arising from pre-existing trends and ensures a more robust estimation of treatment effects.

5 Results

5.1 Summary Statistics

Panel A of Table 1 presents the distribution of tasks exposed to generative AI. The statistics show that 55% of tasks are not exposed to generative AI, 28% are directly exposed (automatable), and 16% are indirectly exposed through LLM-powered applications, with only 0.4% requiring image-based capabilities. This indicates that while generative AI has meaningful reach, exposure is uneven across tasks. Panel B of Table 1 further shows that the automatable-task share averages 0.32, with some occupations reaching as high as 0.86. The automation-exposure diversity index has a mean of 0.30 and is capped at 0.50 by design, indicating variation in how evenly tasks within occupations are split between AI-exposed and non-exposed. Lastly, Panel C highlights the occupations at the extremes of each index. Clerical and transcription roles (e.g., correspondence clerks, medical transcriptionists) have the highest automatable-task shares. In contrast, scientific, medical, and professional roles (e.g., neuropsychologists, medical dosimetrists, and fish and game wardens) top the automation-exposure diversity index. These jobs combine tasks where generative AI can assist (e.g., summarizing notes) with leftover tasks requiring judgment or decision-making, explaining their higher augmentation potential.

Panel A of Table 2 presents summary statistics for the sample used in the automatable-task share analysis, based on firm-occupation quartile-quarter observations. The sample includes occupations in the top quartile of the automatable-task share, along with those in the third and fourth quartiles, which serve as the comparison group, and comprises of 524,050 firm-quartile-quarter observations. On average, firms post 145 job openings per quartile per quarter, while job postings associated with occupations in the top quartile of automatable-task share account for an average of 201 postings. Moreover, job-postings require, on average, 24.4 distinct skills (median = 21.5) across all occupations, compared to 31.1 distinct skills (median = 27.0) among occupations with the highest automatable-task share.

Panel B of Table 2 presents summary statistics for the automation-exposure diversity analysis sample, also based on firm-occupation quartile-quarter observations. This sample includes occupations in the top quartile of the automation-exposure diversity, and those in the third and fourth quartiles, which serve as the comparison group. In total this sample comprises of 574,575 firm-quartile-quarter observations. On average, job postings associated in this set of analyses occur slightly more frequently than the sample in Panel A, averaging 155 postings per quartile per quarter, with 179 postings in the top automation-exposure diversity quartile. The average job postings in this sample also requires 28.8 distinct skills, while those in the top automation-exposure diversity quartile require 27.9 skills. Generative AI-exposed skills are also somewhat less concentrated among high automation-exposure diversity occupations, averaging 15.5 generative AI-exposed skills compared to 18.0 in the full sample. New skill emergence is also modestly lower in these occupations, averaging 8.5 relative to 9 new skills.

5.2 Main Results: Generative AI's Impact on Labor Demand and Skills

Next, we study the heterogeneous impact of generative AI on the labor market. Table 3 reports SDID estimates at the firm-exposure quartile-quarter level, where each observation aggregates all postings for a firm, in a given quarter, across occupations in the same exposure quartile.

Columns (1)-(2) present results for the top quartile of the automatable-task share, and Columns (3)-(4) for the top quartile of automation-exposure diversity. On average, firms reduce postings in high automatable-task-share quartiles by 26.4 jobs per quarter relative to controls, while increasing postings in high automation-exposure-diversity quartiles by 28.3 jobs per quarter relative to controls.[12] The corresponding Poisson DiD models (by applying SDID-derived synthetic weights) imply a 13% decline and a 20% increase, respectively, relative to controls.

The decline in job postings for high labor-substitution potential occupations likely reflects generative AI's ability to perform structured cognitive tasks more efficiently than humans. As firms adopt generative AI, they reduce reliance on labor from high exposed occupations, resulting in lower demand. This mechanism aligns with prior research on capital-augmenting technologies, which reduce labor demand by automating human-performed tasks (e.g., Acemoglu & Autor, 2011; Acemoglu & Restrepo, 2018a; Autor et al., 2003).

In contrast, the increase in job postings for occupations with high labor-augmentation potential supports the view that generative AI enhances productivity in these roles. These occupations consist of a mix of tasks: some of which can be automated, while others remain unaffected by AI. As the automatable tasks are delegated to AI, workers may have more time to focus on judgment-intensive or creative tasks, potentially raising worker productivity and increasing job demand. These results are consistent with H1 and suggests that generative AI reshapes labor demand heterogeneously across the occupational exposure distribution.

Notably, these patterns align with canonical task-based models of technological change which highlight technology's substitution and complementary effects on labor (Acemoglu and Restrepo, 2018a; 2019). Autor et al. (2003) show that computerization substitutes for routine cognitive tasks while complementing non-routine analytic and interactive ones. In a similar

[12] Figure S1 shows dynamic SDID event-study estimates. Both treatment groups exhibit flat pre-trends prior to the event and post-treatment dynamics confirm the main SDID results reported in Table 4.

vein, our findings show that generative AI reduces demand for occupations with homogenous, highly automatable task structures (analogous to routine cognitive work), while increasing demand for occupations with a heterogeneous mix of automatable and less automatable tasks.

We next analyze how generative AI affects the specific skills required for these occupations (Table 4). If automation is reducing hiring demand by substituting for structured cognitive tasks, we expect a corresponding decline in required skills in the top quartile of the automatable-task share. Conversely, if generative AI augments productivity in other occupations, we expect an increase in skill demand, particularly for capabilities that enable workers to integrate generative AI effectively, among occupations in the top quartile of automation-exposure diversity.

Panel A of Table 4 confirms the substitution mechanism: occupations in the top quartile of the automatable-task share show significant declines in required skills following ChatGPT's release. SDID estimates indicate an average drop of 1.41 required skills per firm-occupation-quartile-quarter, with generative AI-exposed skills falling by 0.78 and new skills by 0.45. Poisson DiD estimates (based on SDID-derived synthetic weights) translate to percentage declines of 6.7%, 6.6%, and 6.7%, respectively. These reductions suggest that generative AI simplifies job content by absorbing not only specific generative AI-exposed tasks but also reducing the scope for specialized and newly emerging skills. The decline in new skills indicates fewer opportunities for task innovation as these occupations become increasingly defined by functions that generative AI can directly execute.

Panel B of Table 4 presents the results for occupations in the top quartile of automation-exposure diversity. In contrast to the automation results, we find significant increases across all three skill measures. The SDID estimates show that the average number of required skills per firm-occupation quartile-quarter increases by 2.02, generative AI-exposed skills by 1.41, and new skills by 0.44. The Poisson DiD models further show estimated percentage increases

of 7.1%, 8.2%, and 5.5%, respectively. These patterns are consistent with the augmentation impact of generative AI, where technology complements rather than replaces human labor. By absorbing portions of task execution, generative AI enables workers to shift effort toward judgment-intensive, creative, and collaborative activities, leading to broader and more complex skill requirements. The observed increase in new skills suggests that generative AI is not only enhancing productivity but also expanding the frontier of workplace capabilities, consistent with Autor et al. (2024)'s view of technology as a complementary force for innovation.

Overall, our findings on the changes of skills requirements provide strong support for Hypotheses 2 and 3, by demonstrating that the introduction of generative AI leads to divergent shifts in skill demand across occupations in the top quartile of the automatable-task share and automation-exposure diversity. Consistent with theories of technological change, occupations with high automatable-task shares show simplified and stagnant skill requirements, while those with high automation-exposure diversity exhibit broader skill sets and greater task innovation. Taken together, these results indicate a broader shift in the composition of labor demand toward more skill-intensive tasks in areas where human-AI complementarity is strongest, and away from roles where work complexity and specialized human input can be fully automated. These results also align with the broader literature on skill-biased technical change, which argues that technological advances increase the productivity and demand for skilled workers relative to less skilled ones (Katz & Murphy, 1992; Autor et al., 2003).

5.2.1 Heterogeneity by organizational form

Building on the main analysis, we next examine the workforce impact of generative AI across organizations. Organizational characteristics such as size, sector, and institutional constraints may shape both the pace of AI adoption and the ways it affects hiring decisions. For instance, larger and public firms may be more capable of integrating generative AI into their workflows, while smaller or government organizations may face budgetary or regulatory barriers.

To explore these differences, we estimate two specifications. First, we run Poisson DiD models separately by organizational type, applying SDID-derived synthetic weights to balance pre-treatment dynamics. These models yield directly interpretable percentage changes in job postings and are summarized in Figure 2. Second, we report SDID estimates in Table 5, which provide level effects for each organizational type using directly matched synthetic controls.

The Poisson DiD estimates in Figure 2 suggest substantial heterogeneity. For occupations in the top quartile of the automatable-task share, job postings decline significantly by 15.8 percent in large firms and 9.6 percent in public firms relative to control quartiles. Smaller effects are observed for small firms (-8.0 percent), while government and non-profit organizations show small, positive but insignificant changes. Table 5 corroborates these findings with SDID estimates for large and small firms, but yields insignificant changes for public, government, and non-profit organizations. This modest discrepancy for public firms could stem from differences in estimator structure and sensitivity to sample size. SDID constructs matched synthetic controls for stability, while the Poisson DiD estimator applies SDID-derived weights within a multiplicative model and can be more sensitive to skewness and sparsity.

In contrast, for occupations in the top quartile of the automation-exposure diversity, job postings increase significantly by 28.3 percent in large firms and 22.1 percent in public firms relative to control quartiles, indicating that these organizations are at the forefront of generative-AI-enabled workforce expansion. The effects are smaller and not statistically significant for small firms (+7.8 percent), non-profits (+3.2 percent), and government organizations (-5.7 percent). Corresponding SDID estimates reported in Panel B of Table 5 confirm positive effects for large and public firms, with insignificant effects for government, small, and non-profit organizations. Despite these variations, the general pattern of stronger responses in larger, better-resourced organizations holds across specifications.

Overall, these results suggest that large and public firms appear to be adopting dual strategies, augmenting tasks in occupations with high automation-exposure diversity while reducing postings in occupations with high automatable-task share. On the other hand, smaller firms and government organizations show weaker or inconsistent responses to generative AI. These results therefore suggest that organizational capacity and sectoral constraints play a role in shaping the pace and scope of generative AI-related labor market shifts.

5.2.2 Heterogeneity by industry sector

Next, we examine the heterogeneity in generative AI's workforce impact across industry sectors, recognizing that task structure, exposure to generative AI, and investment incentives vary across the economy. Prior work (e.g., Eisfeldt et al., 2023) show that finance and technology sectors as the most exposed to generative AI. Accordingly, we group firms into three broad sectors based on NAICS-2 codes: Finance (NAICS 52), Tech (NAICS 51 and 54), and Others (all remaining sectors). We estimate Poisson DiD models weighted by SDID-derived synthetic weights using firm-occupation quartile-quarter data (percentage changes; Figure 3) and report the corresponding SDID estimates in Table 6.

Across both Figure 3 and Table 6, we find that job postings decline for occupations in the top quartile of the automatable-task share across all sectors, with the largest reductions in Finance (34%), Tech (26%), followed by firms in the rest of the economy (11%).[13] For occupations in the top quartile of the automation-exposure diversity, we find that job postings increase in all sectors. However, on a percentage change basis, estimated using the weighted Poisson DiD model (Figure 3), we find a statistically significant 20% increase in the job

[13] While the direction and relative magnitude of the effects are broadly consistent across the two specifications, the SDID estimates show stronger statistical significance than the SDID-weighted Poisson DiD models as shown in Figure 3. This divergence likely stems from methodological differences as discussed in the previous section.

postings of the firms in the sectors outside of Finance and Tech, while the firms in these two sectors exhibit statistically insignificant changes in job postings.

Taken together, we find that firms in the Finance and Tech sectors exhibit relatively larger decrease in job demand for high automatable-task-share occupations but exhibit relatively lower increase in job demand for high task-diversity occupations. These results suggest that in these information-intensive industries, generative AI is more likely used to replace labor than to complement it, perhaps in structured cognitive tasks like compliance review, fraud detection, or software development. In contrast, firms in the other sectors exhibit higher job posting increases in high automation-exposure diversity occupations, consistent with broader complementarity between generative AI and human input in less structured workflows. While these patterns provide early descriptive evidence on the heterogeneity in generative AI's labor impact across sectors, we note that these patterns should be interpreted cautiously and not as definitive evidence on firm-level deployment motives.

5.3 Characteristics of Unexposed Tasks

To deepen our understanding of the tasks that remain resistant to generative AI, we turn next from descriptive statistics to task-level analyses. Specifically, we study the skills characteristics of tasks that remain unexposed in occupations with high augmentation potential, by estimating linear probability models that regress an indicator for whether a task is unexposed (E0), on the shares of different skill groups relative to basic skills.[14]

Table 7 shows that tasks with higher social and technical skill shares are significantly more likely to remain unexposed to generative AI relative to baseline basic skills in the full

[14] Skill shares are computed from *O*NET* task-skill linkages as the fraction of skills within each task that belong to a given skill group. To avoid perfect collinearity, we omit Basic Content and Basic Process skill sets as baseline categories, so all coefficients are interpreted relative to these omitted groups.

sample. Unexposed tasks are more likely to comprise resource management skills also load positively, less likely to be systems skills, while exhibiting no statistical relation with complex problem-solving skills. Moreover, we examine these patterns among the occupations in the top quartile of automation-share diversity, and the interaction terms indicate that unexposed tasks comprise of more social and technical skills compared to other occupations. These results highlight that the interpersonal and hands-on technical elements of work are not replaceable by generative AI, especially in settings where augmentation potential is high.[15]

5.4 Robustness Analysis

Lastly, we assess the robustness of our main findings using a range of alternative specifications and subsamples, addressing potential concerns related to aggregation bias, pre-trend violations, and unobserved firm-level shocks.

First, we address concerns relating to aggregation bias. To ensure that our results are not an artifact of aggregating data at the firm-occupation quartile-quarter level., we collapse the data and re-estimate the models at the occupation-quartile-quarter level. This specification strips away firm-level heterogeneity, focusing instead on whether results reflect aggregate exposure patterns across occupations. Treatment is again defined as the top quartile of the automatable-task share or automation-exposure diversity, compared to the third and fourth quartiles. As shown in Table S2, the estimates remain strong and statistically significant: job postings decline by 1,851 per occupation quartile-quarter for occupations in the top quartile of the automatable-task share and rise by 1,211 for those in the top quartile of the automation-

[15] For context the top social skills required are: (1) Maintain and develop positive business relations, (2) Build and maintain partnerships with federal agencies, (3) Develop liaisons with external organizations. While the top technical skills required are: (1) Lift and move raw materials or finished products. (2) Drive and operate mobile medical units, (3) Operate machines to blend and extrude clay.

exposure diversity. Poisson regressions yield percentage changes of -22.5% and +19.7%, respectively, closely matching our main findings.

Second, we address potential endogeneity concerns related to industry-specific shocks. Our main SDID design mitigates concerns about parallel trends by optimally re-weighting control units, but it does not explicitly absorb sector-level shocks that could simultaneously influence hiring. To test robustness, we re-estimate a conventional DiD at the finer firm-occupation-quarter level (rather than the firm-exposure quartile-quarter level used in the main analysis). This alternative unit of analysis allows us to both (1) strip away the aggregation to quartiles and (2) directly include industry-by-quarter fixed effects to absorb time-varying shocks at the NAICS-6 sector level.

Results reported in Table S3 show that the main findings are robust across these specifications. For occupations in the top quartile of the automatable-task share (Panel A), Poisson estimates indicate percentage declines of 12.8% with quarter FE and 12.1% with industry-by-quarter FE. For occupations in the top quartile of the automation-exposure diversity (Panel B), we find percentage increases of 15.6% and 18.4%, respectively. Notably, the fact that adding industry-by-quarter fixed effects leaves the estimates largely unchanged suggests that our results are not driven by contemporaneous sector-level shocks but rather reflect differences in occupational exposure to generative AI.

Third, we address endogeneity concerns relating to the systematic differences between treatment and control units. To address this, we implement a matched DiD approach at the firm-occupation-quarter level that improves covariate balance prior to estimation.[16] Table S4

[16] Specifically, we perform a two-step procedure to match treatment and control units. First, we perform exact matching on three key structural characteristics: (1) Industry, defined at the NAICS-2 level; (2) Firm size, using LightCast's employee count group variable, which categorizes firms into eight groups based on their number of employees; and (3) Job Zone, an *O*NET*-based classification that groups occupations into five categories based

presents matched difference-in-differences estimates for job postings. For occupations in the top quartile of the automatable-task share (Panel A), the Poisson estimates suggest percentage declines of 19.4% to 17.0%, depending on the specification. For occupations in the top quartile of the automation-exposure diversity (Panel B), the Poisson estimates indicate increases of 13.0% to 11.9%.[17] Overall, these results remain directionally consistent with main analysis and further reinforce that the observed effects are not driven by covariate imbalance or baseline differences between treated and control groups.

Fourth, we address the concern that the main results are driven by changes in the membership of firms within treatment and control groups over time. We address this by restricting the sample to firms that post both treated and control occupations during the study period, which ensures that estimated treatment effects are identified using only within-firm variation. Table S5 reports SDID and weighted Poisson DiD estimates at the firm-exposure quartile-quarter level based on this restricted sample.[18] The results remain nearly identical to

---

on the typical education, experience, and training required to perform the job. Second, within each exact match group, we apply propensity score matching with replacement, using the two nearest neighbors based on the total number of firm-occupation job postings in the pre-treatment period.

[17] To visualize the dynamics of the treatment effect, Figure S2 presents event study estimates from matched Poisson DiD models for occupations in the top quartile of the automatable-task share and those in the top quartile of the automation-exposure diversity. While the post-treatment effects align closely with our main findings, we observe some deviations from parallel pre-trends in both groups, particularly in the quarters immediately preceding the introduction of ChatGPT. These limited pre-trends, despite improved covariate balance through matching, underscores the value of the SDID approach, which re-weights control observations to closely match the treated group's pre-treatment trajectory. Thus, while the matched DiD results provide supportive evidence, we continue to rely on SDID as our preferred specification for estimating the causal impact of generative AI exposure.

[18] Notably, the number of observations in the restricted sample is only slightly smaller than in the full sample, indicating that most firms post both treated and control occupations. This implies that our main results are already

those from the full sample: for automatable-task share occupations, the Poisson estimate indicates a 12.7% decline (vs. 12.9%); for automation-exposure diversity occupations, the Poisson estimates suggests a 20.5% increase (vs. 20.2%).

Fifth, we address measurement concerns that our findings could be sensitive to the definition of automatable tasks. Following Eloundou et al (2024), our baseline specification defines the automatable-task share as the importance-weighted average of task exposure scores with weights 1 for E1 (direct LLM) and 0.5 for E2 (LLM-powered apps), and 0 for E3. To test sensitivity, we check robustness for the automatable-task share defined in two ways: (i) E1+E2 (treats both direct LLM and LLM-app tasks as fully automatable) and (ii) E1+E2+E3 (adds image-dependent tasks). Analysis with these specifications, reported in Table S6, shows Poisson estimates that imply 13 and15% lower postings, respectively. In comparison, these effects are only slightly larger in magnitude than the baseline Poisson estimate, indicating that our main results do not hinge on whether E2 (and E3) tasks are counted as fully automatable.

Sixth, we perform additional robustness analysis of our main results with alternative indices that proxy for labor-substitution and labor-augmentation potential. Following Handa et al (2025), we construct occupation-level of both types of indices from the Anthropic Economic Index (AEI) by mapping AEI's automation and augmentation task records to *O*NET* task statements and aggregating to occupations using *O*NET* "Importance" weights (importance-weighted means).[19] Table S7 reports results for the firm-exposure quartile-quarter

largely identified by within-firm comparisons. The consistency of treatment effects across these specifications further reinforces the robustness of our findings and the credibility of the SDID framework for estimating the causal impact of generative AI exposure.

[19] To ensure usable signal, we retain only tasks with "AEI mass" ≥ 0.75, where AEI mass denotes the share of a task's recorded interactions that are valid (non-filtered). We further note that AEI coverage is limited, and spans

sample, after re-binning occupations into quartiles using AEI-based automation and augmentation scores (top quartile vs. Q3-Q4). For the automation sample, the SDID-weighted Poisson estimate is -6% but is statistically insignificant. For the augmentation sample, the SDID-weighted Poisson estimate is 13% and is significant at the 1% level. These findings suggest that the main results on augmentation effects are positive and robust. On the other hand, automation effects under the AEI specifications attenuate toward zero, plausibly due to AEI's limited task coverage and selection (few mapped tasks per occupation).

Taken together, the robustness analyses across multiple estimation strategies, levels of aggregation, sample restriction, and index definition further confirms that the introduction of ChatGPT has led to a reallocation of labor demand across occupations, with declines concentrated in those in the top quartile of the automatable-task share and gains concentrated in those in the top quartile of the automation-exposure diversity.

5.5 Limitations

We acknowledge several limitations of our study. First, our estimates capture short-run effects on the labor market, and the long-term impacts remain uncertain as generative AI adoption scales. Many firms are still at the early stages of adoption, and thus extrapolating from short-run trends in job postings to long-run impact on employment or wages should be done cautiously. Nonetheless, our evidence consistently points to declines in occupations with high automatable-task share and gains in those with high task-diversity, consistent with theories on the labor-substitution and labor-augmentation potential of technology.

Second, measurement challenges remain. Job postings are used as a proxy for labor demand, but job postings do not always translate into hires (e.g., "ghost" postings; Hao &

only ~20% of *O*NET* tasks, which drops many occupations. Even among included occupations, the mean fraction of tasks covered is ~6% (max 73%), which implies substantial truncation and potential measurement errors.

Simon, 2023) and may reflect signaling rather than genuine intent to hire. We mitigate this concern with within-firm comparisons, matched samples, industry×quarter fixed effects, and robustness at the occupation-quarter level. Moreover, while our indices of automatable-task share and automation-exposure diversity are designed to capture distinct substitution and augmentation channels, some conflation errors may persist. In addition, model capabilities evolve, so some gradual biases in classification may arise due to the stationary assumption that we impose on exposure classification. Finally, quartile cutoffs, while transparent, may obscure marginal cases; on this front, we note that our results are robust to alternative specifications, but future work using continuous measures could provide finer granularity.

Third, our analysis focuses on occupation-level exposure to generative AI and abstracts away from potential heterogeneity in the intensity and timing of generative AI adoption across firms. Our research design defines the treatment exposure to generative AI, as constant across firms and determined prior to ChatGPT's release, which avoids concerns about endogenous firm-level adoption decisions. In reality, some firms may have adopted generative AI tools aggressively, while others may have delayed or avoided adoption altogether, leading to across-firm heterogeneity that may be important to understand or attenuate our main estimates. While firm-occupation and firm-quartile fixed effects help mitigate these concerns, we acknowledge the limitation and believe that it underscores the need for future research on firm-level adoption intensity and its interaction with occupational exposure.

Fourth, while our identification strategy incorporates a broad set of design-based and statistical controls, we cannot fully rule out the presence of unobserved confounders that differentially affect treated and control occupations. For example, firms may have adopted internal productivity tools, restructured teams, or adjusted hiring practices for reasons unrelated to generative AI but correlated with exposure. Moreover, ChatGPT's launch coincided with broader interest in AI and digital technology adoption, which may have differentially affected

certain sectors or occupations. We mitigate these concerns through rich fixed effects and design checks but acknowledge this as a limitation of any non-experimental design.

Fifth, we acknowledge that external validity could be limited with our analysis. Our study focuses on the U.S. labor market, and effects may differ with institutions, adoption costs, and occupational mix elsewhere.

6 Conclusion

We study the labor-market impact of generative AI by exploiting the November 2022 release of ChatGPT and separating occupations along two dimensions, automatable-task share, which maps to labor-substitution potential, and automation-exposure diversity, which maps to labor-augmentation potential of generative AI at the occupation level. Occupations with high automatable-task shares see reduced hiring, while occupations with high task diversity see increased hiring. Skill demands mirror this pattern, with declines (total, generative AI-exposed, and new skills) in the former and increases in the latter. These findings suggest that generative AI reduces labor demand where core tasks are readily automatable but raises demand and broadens required capabilities where automatable and non-automatable tasks co-exist.

Our identification relies on a plausibly exogenous timing shock, and findings are robust across SDID, conventional and matched DiD, within-firm comparisons and alternative index constructions. We also assess heterogeneity across organizational form and industry sectors. Labor demand changes are stronger in larger firms, with declines concentrated in occupations with high automatable-task shares (substitution potential) in Finance and Tech and increases most evident in occupations with high automation-exposure diversity (augmentation potential) in other sectors.

These findings carry both policy and managerial implications. On the policy side, reallocation pressures in high automatable-task-share occupations highlight the need for targeted transition support, such as retraining and reemployment programs. On the managerial side, our task-level analysis shows that the activities most resistant to generative AI in

augmentation-heavy occupations disproportionately rely on social interaction and hands-on technical skills. Combined with the hiring gains observed in high automation-task diversity occupations, this suggests that the value of generative AI lies not only in automating structured tasks but also in amplifying human strengths in areas generative AI has yet to reach.

Acknowledgements and Disclosures

This research was supported by the Digital, Data, and Design Institute at Harvard Business School. We have no financial relationships or other potential conflicts of interest to disclose.

References

Acemoglu, D. (2002a). Directed technical change. *Review of Economic Studies*, 69(4), 781–809. doi:10.1111/1467-937X.00226

Acemoglu, D. (2002b). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1), 7–72. doi:10.1257/0022051026976

Acemoglu, D. (2003). Labor- and capital-augmenting technical change. *Journal of the European Economic Association*, 1(1), 1–37. doi:10.1162/154247603322256756

Acemoglu, D. (2024). The simple macroeconomics of AI. NBER Working Paper No. 32487. National Bureau of Economic Research. doi:10.3386/w32487

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics*, Vol. 4B, pp. 1043–1171. Amsterdam: Elsevier. doi:10.1016/S0169-7218(11)02410-5

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. doi:10.1086/718327

Acemoglu, D., Kong, F., & Restrepo, P. (2024). Tasks at work: Comparative advantage, technology and labor demand. NBER Working Paper No. 32872. National Bureau of Economic Research. doi:10.3386/w32872

Acemoglu, D., & Restrepo, P. (2018a). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans & A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, pp. 197–236. Chicago: University of Chicago Press.

Acemoglu, D., & Restrepo, P. (2018b). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6), 1488–1542. doi:10.1257/aer.20160696

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30. doi:10.1257/jep.33.2.3

Acemoglu, D., & Restrepo, P. (2020). The wrong kind of AI? Artificial intelligence and the future of labour demand. *Cambridge Journal of Regions, Economy and Society*, 13(1), 25–35. doi:10.1093/cjres/rsz022

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, 90(5), 1973–2016. doi:10.3982/ecta19815

Agrawal, A., Gans, J., & Goldfarb, A. (2023a). Similarities and differences in the adoption of general purpose technologies. NBER Working Paper No. 30976. National Bureau of Economic Research. doi:10.3386/w30976

Agrawal, A., Gans, J. S., & Goldfarb, A. (2023b). Do we want less automation? AI may provide a path to decrease inequality. *Science*, 381(6654), 155–158. doi:10.1126/science.adh9429

Alekseeva, L., Azar, J., Giné, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, 71, 102002. doi:10.1016/j.labeco.2021.102002

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118. doi:10.1257/aer.20190159

Autor, D., Chin, C., Salomons, A., & Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. *Quarterly Journal of Economics*, 139(3), 1399–1465. doi:10.1093/qje/qjae008

Autor, D. H., Katz, L. F., Kearney, M. S., Berman, E., & Chandra, A. (2006). The polarization of the U.S. labor market. *American Economic Review*, 96(2), 189–194. doi:10.1257/000282806777212620

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279–1333. doi:10.1162/003355303322552801

Berger, P. G., Cai, W., Qiu, L., & Shen, C. X. (2024). Employer and employee responses to generative AI: Early evidence. SSRN Working Paper. doi:10.2139/ssrn.4874061

Bessen, J. E. (2018). AI and jobs: The role of demand. NBER Working Paper No. 24235. National Bureau of Economic Research. doi:10.3386/w24235

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER Working Paper No. 31161. National Bureau of Economic Research. doi:10.3386/w31161

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications: Profound change is coming, but roles for humans remain. *Science*, 358(6370), 1530–1534. doi:10.1126/science.aap8062

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108, 43–47. doi:10.1257/pandp.20181019

Cao, R., & Iansiti, M. (2022). Digital transformation, data architecture, and legacy systems. *Journal of Digital Economy*, 1(1), 1–19.

Choi, J. H., & Xie, C. (2025). Human + AI in accounting: Early evidence from the field. SSRN Working Paper. doi:10.2139/ssrn.5240924

Clarke, D., Pailañir, D., Athey, S., Imbens, G., & Wager, S. (2023). Synthetic difference-in-differences estimation. arXiv preprint arXiv:2301.11859.

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2), 529–551. doi:10.1016/j.jfineco.2022.08.004

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), eadn5290. doi:10.1126/sciadv.adn5290

Eisfeldt, A. L., Schubert, G., Zhang, M. B., & Taska, B. (2023). The labor impact of generative AI on firm values. SSRN Working Paper. doi:10.2139/ssrn.4436627

Ellingrud, K., Sanghvi, S., Dandona, G. S., Madgavkar, A., Chui, M., White, O., & Hasebe, P. (2023). Generative AI and the future of work in America. McKinsey Global Institute Report.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, 384(6702), 1306–1308. doi:10.1126/science.adj0998

Felten, E. W., Raj, M., & Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Papers and Proceedings*, 108, 54–57.

Felten, E. W., Raj, M., & Seamans, R. (2023). Occupational heterogeneity in exposure to generative AI. SSRN Working Paper. doi:10.2139/ssrn.4414065

Fossen, F. M., Samaan, D., & Sorgner, A. (2022). How are patented AI, software and robot technologies related to wage changes in the United States? *Frontiers in Artificial Intelligence*, 5, 869282. doi:10.3389/frai.2022.869282

Frank, M. R., Ahn, Y.-Y., & Moro, E. (2025). AI exposure predicts unemployment risk: A new approach to technology-driven job loss. *PNAS Nexus*, 4(4), pgaf107. doi:10.1093/pnasnexus/pgaf107

Green, A. (2024). Artificial intelligence and the changing demand for skills in the labour market. OECD Report.

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., … & Ganguli, D. (2025). Which economic tasks are performed with AI? Evidence from millions of Claude conversations. arXiv preprint arXiv:2503.04761.

Hao, S., & Simon, L. (2023, October). Ghost job postings. Revelio Labs Blog.

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *American Economic Review*, 108(7), 1737–1772.

Hui, X., Reshef, O., & Zhou, L. (2024). The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market. *Organization Science*, 35(6), 1977–1989. doi:10.1287/orsc.2023.18441

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics*, 107(1), 35–78.

Liu, J., Foster, M. G., Nan, X., Li, Y., & Tan, Y. (2023). Generate the future of work through AI: Empirical evidence from online labor markets. SSRN Working Paper. doi:10.2139/ssrn.4529739

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

Webb, M. (2019). The impact of artificial intelligence on the labor market. Working Paper, Stanford University.

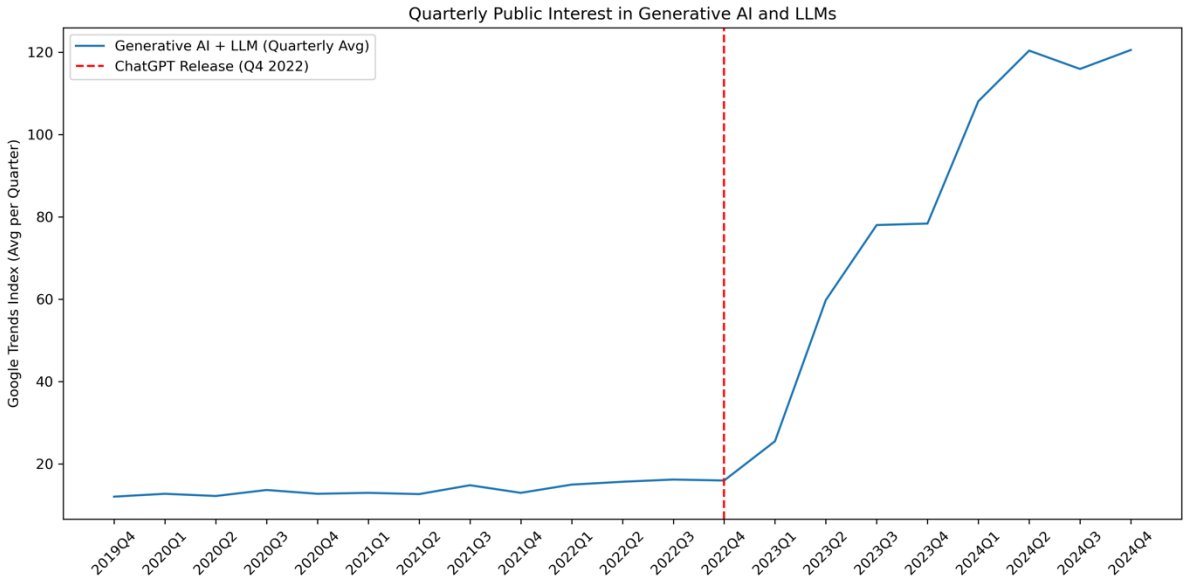
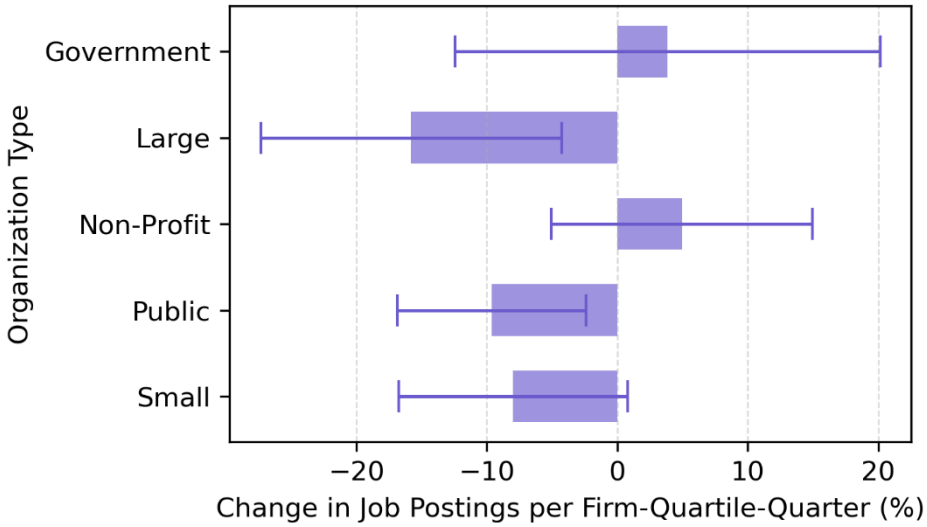
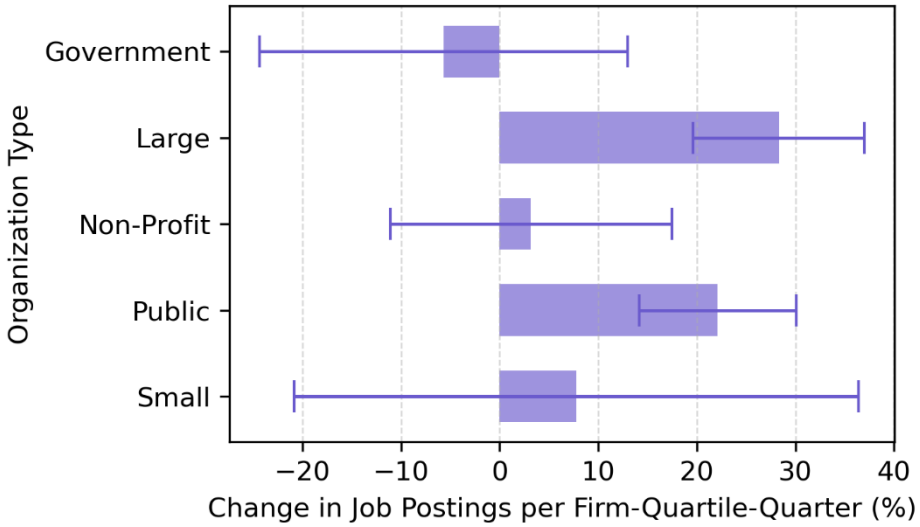


Figure 1. Public Interest in Generative AI and LLMs.
This figure reports the quarterly average Google Trends index for generative AI and large language models (LLMs).



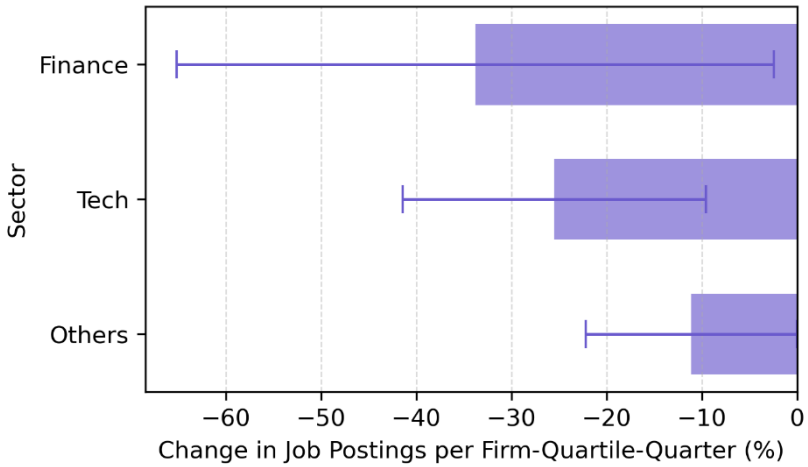
(A) Automation- top quartile of automatable-task share



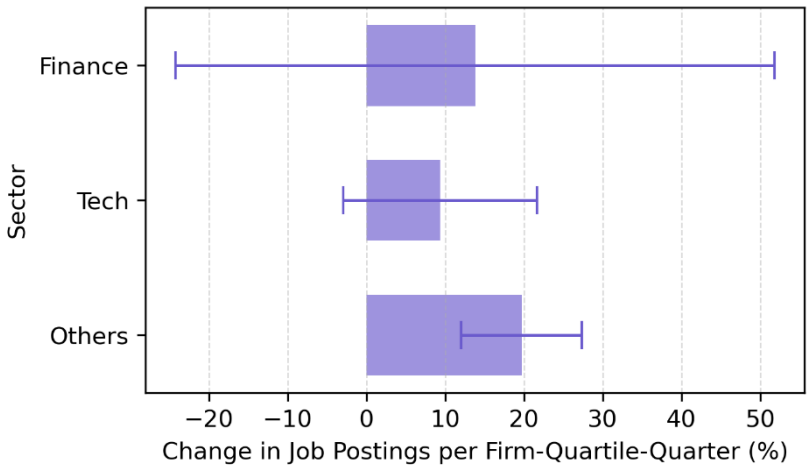
(B) Augmentation- top quartile of automation-exposure diversity

Figure 2. Percentage Change in Job Postings by Firm Type.

Panel (A) displays the estimated percentage change in job postings for occupations in the top quartile of the automatable-task share across firm types following the introduction of ChatGPT. Panel (B) presents the corresponding estimates for occupations in the top quartile of automation-exposure diversity. Estimates are obtained from Poisson DiD models weighted by SDID-derived synthetic weights, comparing top quartile occupations to those in the third and fourth quartiles. Vertical lines represent 95% confidence intervals. Standard errors are clustered at the firm-occupation level.



(A) Automation- top quartile of automatable-task share



(B) Augmentation- top quartile of automation-exposure diversity

Figure 3. Percentage Change in Job Postings by Industry Sector.

Panel (A) displays the estimated percentage change in job postings for occupations in the top quartile of the automatable-task share across industry sectors following the introduction of ChatGPT. Panel (B) presents the corresponding estimates for occupations in the top quartile of the automation-exposure diversity. Estimates are obtained from Poisson difference-in-differences models weighted by SDID-derived synthetic weights, using firm-quartile-quarter data. The models compare occupations in the top quartile of the respective index to those in the third and fourth quartiles. Industry categories are based on NAICS-2 codes: Tech includes NAICS 51 (Information) and 54 (Professional, Scientific, and Technical Services); Finance includes NAICS 52 (Finance and Insurance); all other industries are grouped into Other. Vertical lines represent 95% confidence intervals, and standard errors are clustered at the firm-quartile level.

Table 1. Task Exposure Classifications and Occupational Exposure Indices.

Panel A presents the distribution of task-level exposure classifications to generative AI, as determined by GPT-4o. Tasks are grouped into four categories: E0 (no exposure), E1 (direct exposure), E2 (exposure through LLM-powered applications), and E3 (exposure through image capabilities). Panel B summarizes the distribution of the automatable-task share and automation-exposure diversity across all occupations. The automatable-task share reflects the proportion of occupational time spent on tasks that can be fully executed by an LLM, while the automation-exposure diversity captures the weighted heterogeneity in generative AI-exposed versus non-exposed tasks. Panel C lists the top five occupations with the highest values of each index, highlighting those most prone to substitution and those most conducive to augmentation.

Panel A: Task classification distribution (GPT-4o classifications)

Exposure (Automation)	Task Count	Task Share
No exposure (E0)	10,611	55.08%
Direct exposure (E1)	5458	28.33%
Exposure by LLM-powered applications (E2)	3111	16.15%
Exposure given image capabilities (E3)	85	0.44%
Total	19,265	100%

Panel B: Summary statistics for occupational exposure indices.

Index	Number of Observations	Mean	Standard deviation	Median	p75	Max
Automatable-task share	923	0.316	0.157	0.299	0.454	0.860
Automation-exposure diversity	923	0.298	0.197	0.357	0.459	0.500

Panel C: Top occupations by task automation and augmentation potential.

Occupation	Automatable-task share	Occupation	Automation-exposure diversity
Correspondence Clerks	0.859608	Clinical Neuropsychologists	0.500000
Interpreters and Translators	0.797865	Medical Dosimetrists	0.500000
Court, Municipal, and License Clerks	0.761533	Fish and Game Wardens	0.499999
Medical Transcriptionists	0.742068	Agricultural Engineers	0.499993
Telemarketers	0.734706	Cartographers and Photogrammetrists	0.499991

Table 2. Summary Statistics

This table reports summary statistics for firm-occupation quartile-quarter observations used in the automation and augmentation analyses. Panel A includes job postings and skill measures for occupations in the top quartile of the automatable-task share, compared to those in the third and fourth quartiles, which serve as the comparison group. Panel B presents the corresponding statistics for the task-diversity sample. "Total job postings" refers to the number of job postings associated with a given firm-occupation quartile in a given quarter. "Job postings in top quartile" refers specifically to those associated with occupations in the top quartile of the relevant index. Skill measures include the average number of required skills, generative AI-exposed skills, and newly emerging skills per firm-occupation quartile-quarter.

Panel A. Automatable-task share sample

Variable	Observation	Mean	Median	Std. Dev.	Min	Max
Job Postings (per firm-quartile- quarter)						
Total job postings	524,050	144.99	28	822.93	0	148,169
Job postings in top quartile of automatable-task share	193,750	200.92	48	1,072.08	0	148,169
Skill Measures (per firm-quartile-quarter)						
Total required skills	524,050	24.36	21.47	16.88	0	746.29
Total required skills in top quartile of automatable-task share	193,750	31.08	27.00	20.29	0	746.29
Generative AI-Exposed Skills (per firm-quartile-quarter)						
Generative AI-exposed skills	524,050	14.15	11	12.98	0	567.53
Generative AI-exposed skills in top quartile of automatable-task share	193,750	21.87	18.56	15.89	0	576.53
New Skills (per firm-quartile-quarter)						
New skills	524,050	7.76	6.75	6.65	0	259.00
New skills in top quartile of automatable-task share	193,750	10.23	9.32	6.46	0	173.19

Panel B. Automation-exposure diversity sample

Variable	Observation	Mean	Median	Std. Dev.	Min	Max
Job Postings (per firm-quartile- quarter)						
Total job postings	574,575	155.18	32.00	732.06	0	111,483
Job postings in top quartile of automation-exposure diversity	191,450	179.27	31.00	772.34	0	38,175
Skill Measures (per firm-quartile-quarter)						
Total required skills	574,575	28.79	25.08	18.96	0	530.58
Total required skills in top quartile of automation-exposure diversity	191,450	27.92	25.45	15.85	0	343.00
Generative AI-Exposed Skills (per firm-quartile-quarter)						
Generative AI-exposed skills	574,575	18.01	14.75	14.62	0	390.11
Generative AI-exposed skills in top quartile of automation-exposure diversity	191,450	15.49	13.50	10.47	0	301.00
New Skills (per firm-quartile-quarter)						
New skills	574,575	9.05	8.00	7.04	0	243.50
New skills in top quartile of automation-exposure diversity	191,450	8.51	7.42	6.88	0	243.50

Table 3. Impact of ChatGPT's Introduction on Hiring Demand

This table reports the effects of ChatGPT's introduction on hiring demand, as measured by job postings, for occupations in the top quartile of the automatable-task share (Columns 1-2) and automation-exposure diversity (Columns 3-4) relative to those in the third and fourth quartiles. SDID estimates (Columns 1 and 3) use a synthetic control group constructed to match pre-treatment trends. Poisson regression estimates (Columns 2 and 4) apply SDID-derived synthetic weights to balance pre-treatment dynamics; observations with zero weights are dropped, leading to a smaller estimation sample. SDID standard errors are bootstrapped using 1,000 replications, and Poisson regression standard errors are clustered at the firm-quartile exposure level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Automatable-Task Share		Automation-exposure diversity	
	(1) SDID estimate on Job Postings	(2) DiD estimate on Job Postings (Poisson Regression)	(3) SDID estimate on Job Postings	(4) DiD estimate on Job Postings (Poisson Regression)
Top Quartile x Post-ChatGPT	-26.40***	-0.129**	28.33***	0.202***
	(8.965)	(0.0492)	(5.122)	(0.0333)
Observations	524,050	441,200	574,575	574,575

Table 4. Impact of ChatGPT's Introduction on Skills Demand

This table reports the estimated impact of ChatGPT's introduction on skill demand for occupations in the top quartile of the automatable-task share (Panel A) and the automation-exposure diversity (Panel B), compared to those in the third and fourth quartiles. Columns (1), (3), and (5) present SDID estimates using a synthetic control group matched on pre-treatment trends. Columns (2), (4), and (6) report Poisson DiD estimates weighted using SDID-derived synthetic weights; observations with zero weights are excluded from the Poisson models. The dependent variables are the average number of required skills, generative AI-exposed skills, and new skills per firm-occupation quartile-quarter. SDID standard errors are bootstrapped using 1,000 replications; Poisson standard errors are clustered at the firm level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. High automatable-task share occupations

	All Skills		Generative AI-Exposed Skills		New Skills	
	(1) SDID estimate	(2) DiD estimate (Poisson Regression)	(3) SDID estimate	(4) DiD estimate (Poisson Regression)	(5) SDID estimate	(6) DiD estimate (Poisson Regression)
Top Quartile x Post-ChatGPT	-1.407***	-0.0673***	-0.783***	-0.0665***	-0.451***	-0.0674***
	(0.101)	(0.00329)	(0.0737)	(0.00409)	(0.0558)	(0.00517)
Observations	524,050	446,075	524,050	441,325	524,050	443,700

Panel B. High automation-exposure diversity occupations

	All Skills		Generative AI-Exposed Skills		New Skills	
	(1) SDID estimate	(2) DiD estimate (Poisson Regression)	(3) SDID estimate	(4) DiD estimate (Poisson Regression)	(5) SDID estimate	(6) DiD estimate (Poisson Regression)
Top Quartile x Post-ChatGPT	2.024***	0.0715***	1.412***	0.0824***	0.443***	0.0546***
	(0.137)	(0.00328)	(0.0714)	(0.00364)	(0.0408)	(0.00381)
Observations	574,575	443,950	574,575	443,950	574,575	574,575

Table 5. Heterogeneous Impact of ChatGPT's Introduction on Hiring Demand Across Organizational Type

This table reports SDID estimates of the impact of ChatGPT's introduction on hiring demand, as measured by job postings, across firm types. Panel A represents results for occupations in the top quartile of the automatable-task share, and Panel B for those in the top quartile of the automation-exposure diversity. Each estimate compares top-quartile occupations to those in the third and fourth quartiles, within the same firm type, using firm-occupation-quartile-quarter observations. Standard errors, shown in parentheses, are bootstrapped using 1,000 replications. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. High automatable-task share occupations

	(1) Government	(2) Large	(3) Non-profit	(4) Public	(5) Small
Top Quartile x Post-ChatGPT	0.0838	-6.298***	-0.247	-0.483	-1.744***
	(0.510)	(1.448)	(0.271)	(0.480)	(0.495)
Observations	119,800	1,004,675	92,850	609,375	296,100

Panel B. High automation-exposure diversity occupations

	(1) Government	(2) Large	(3) Non-profit	(4) Public	(5) Small
Top Quartile x Post-ChatGPT	-0.0265	14.91***	1.083	4.857***	1.234
	(0.308)	(2.147)	(1.127)	(1.065)	(2.493)
Observations	125,750	1,007,625	93,000	617,400	303,075

Table 6. Heterogeneous Impact of Generative AI on Hiring Demand Across Industry Sectors

This table reports SDID estimates of the impact of ChatGPT's introduction on hiring demand, as measured by job postings, across industry sectors. Panel A presents results for occupations in the top quartile of the automatable-task share, and Panel B for those in the top quartile of the automation-exposure diversity. Each estimate compares occupations in the top quartile of the respective index to those in the third and fourth quartiles, within a given industry, using firm-occupation-quartile-quarter observations. Industry classifications are based on NAICS-2 codes: Finance includes NAICS 52; Tech includes NAICS 51 and 54; all remaining sectors are grouped under Other. The standard errors, shown in parentheses, are bootstrapped using 1,000 replications. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. High automatable-task share occupations

	(1) Finance	(2) Technology	(3) Other
Top Quartile x Post-ChatGPT	-160.0***	-62.83***	-19.69***
	(55.00)	(19.68)	(5.469)
Observations	21,625	69,275	433,150

Panel B. High automation-exposure diversity occupations

	(1) Finance	(2) Technology	(3) Other
Top Quartile x Post-ChatGPT	24.60***	10.22***	32.75***
	(7.878)	(3.282)	(6.722)
Observations	31,875	85,850	456,850

Table 7. Characteristics of Unexposed Tasks in High Automation-Share Diversity Occupations

Panel A reports linear probability models where the dependent variable is an indicator for whether a task is unexposed (E0). Explanatory variables are task-level skill-group shares; basic content and basic process are omitted categories. All models absorb SOC fixed effects and cluster standard errors at the SOC level. Column (1) uses the full sample; column (2) compares occupations in the top quartile of automation-share diversity to those in the 3rd and 4th quartiles (excluding the 2nd quartile). Panel B provides examples of unexposed tasks with especially high social or technical skill shares. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Skill characteristics of unexposed tasks

	(1) Full sample	(2) No Q2 sample
Social skills	0.492***	0.356***
	(0.0297)	(0.0328)
Top task diversity quartile × social skills	0.136**	0.271***
	(0.0633)	(0.0649)
Technical skills	0.474***	0.423***
	(0.0164)	(0.0184)
Top task diversity quartile × technical skills	0.153***	0.204***
	(0.0407)	(0.0416)
Complex problem solving skills	0.0175	-0.00613
	(0.0945)	(0.0976)
Top task diversity quartile × complex problem solving skills	0.209	0.232
	(0.361)	(0.362)
Resource management skills	0.343***	0.303***
	(0.0459)	(0.0560)
Top task diversity quartile × resource management skills	-0.0139	0.0267
	(0.104)	(0.109)
Systems skills	-0.179***	-0.154***
	(0.0552)	(0.0590)
Top task diversity quartile × systems skills	-0.0576	-0.0834
	(0.188)	(0.189)
Constant	0.308***	0.326***
	(0.00596)	(0.00675)
SOC FE	Yes	Yes
Observations	15,706	11,774
R-squared	0.352	0.396

Internet Appendix for Displacement or Complementarity? The Labor Market Impact of Generative AI

Wilbur Xinyuan Chen
Hong Kong University of Science and Technology

Suraj Srinivasan
Harvard Business School
Digital, Data, Design Institute at Harvard

Saleh Zakerinia
Ohio State University

Appendix A: GPT Prompts

GPT prompt for task exposure scores

```
## E0 - No exposure:
Label tasks E0 if none of the above clearly decrease the time it takes for an experienced worker to complete the task with high quality by at least half. Some examples:
- If a task requires a high degree of human interaction (for example, in-person demonstrations) then it should be classified as E0.
- If a task requires precise measurements then it should be classified as E0.
- If a task requires reviewing visuals in detail then it should be classified as E0.
- If a task requires any use of a hand or walking then it should be classified as E0.
- Tools built on top of the LLM cannot make any decisions that might impact human livelihood (e.g.hiring, grading, etc.). If any part of the task involves collecting inputs to make a final decision (as opposed to analyzing data to inform a decision or make a recommendation) then it should be classified as E0. The LLM can make recommendations.
- Even if tools built on top of the LLM can do a task, if using those tools would not save an experienced worker significant time completing the task, then it should be classified as E0.
- The LLM and systems built on top of it cannot do anything that legally requires a human to perform the task.
- If there is existing technology not powered by an LLM that is commonly used and can complete the task then you should mark the task E0 if using an LLM or LLM-powered tool will not further reduce the time to complete the task. When in doubt, you should default to E0.

## E1 - Direct exposure:
Label tasks E1 if direct access to the LLM through an interface like ChatGPT or the OpenAI playground alone can reduce the time it takes to complete the task with equivalent quality by at least half. This includes tasks that can be reduced to:
- Writing and transforming text and code according to complex instructions,
- Providing edits to existing text or code following specifications,
- Writing code that can help perform a task that used to be done by hand,
- Translating text between languages,
- Summarizing medium-length documents,
- Providing feedback on documents,
- Answering questions about a document,
- Generating questions a user might want to ask about a document,
```

- Writing questions for an interview or assessment,
- Writing and responding to emails, including ones that involve refuting information or engaging in a negotiation (but only if the negotiation is via written correspondence),
- Maintain records of written data, - Prepare training materials based on general knowledge, or
- Inform anyone of any information via any written or spoken medium.

E2 - Exposure by LLM-powered applications:
Label tasks E2 if having access to the LLM alone may not reduce the time it takes to complete the task by at least half, but it is easy to imagine additional software that could be developed on top of the LLM that would reduce the time it takes to complete the task by half. This software may include capabilities such as:
- Summarizing documents longer than 2000 words and answering questions about those documents,
- Retrieving up-to-date facts from the Internet and using those facts in combination with the LLM capabilities,
- Searching over an organization's existing knowledge, data, or documents and retreiving information,
- Retrieving highly specialized domain knowledge,
- Make recommendations given data or written input,
- Analyze written information to inform decisions,
- Prepare training materials based on highly specialized knowledge,
- Provide counsel on issues, and
- Maintain complex databases.

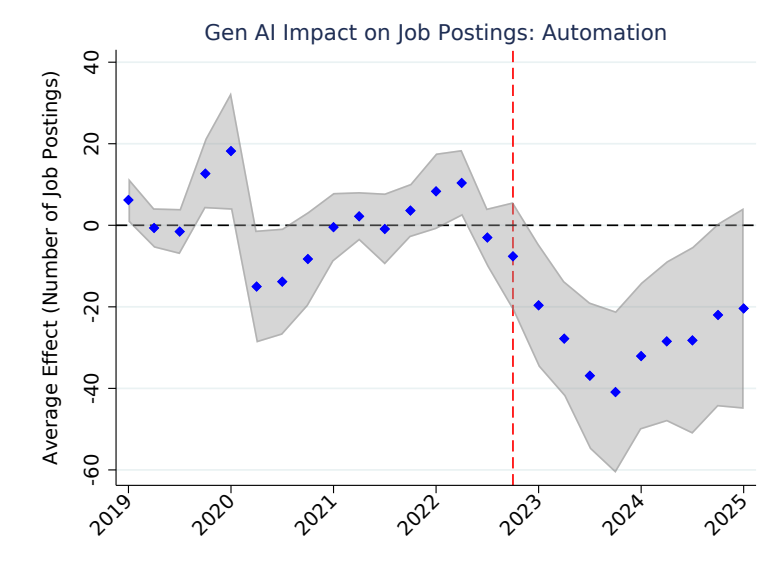
E3 - Exposure given image capabilities:
Suppose you had access to both the LLM and a system that could view, caption, and create images as well as any systems powered by the LLM (those in E2 above). This system cannot take video as an input and it cannot produce video as an output. This system cannot accurately retrieve very detailed information from image inputs, such as measurements of dimensions within an image. Label tasks as E3 if there is a significant reduction in the time it takes to complete the task given access to a LLM and these image capabilities:
- Reading text from PDFs,
- Scanning images, or
- Creating or editing digital images according to instructions. The images can be realistic but they should not be detailed. The model can identify objects in the image but not relationships between those options

GPT prompt for skill exposure score
You are an expert in AI and technology skill analysis with a deep understanding of generative AI and its applications across various

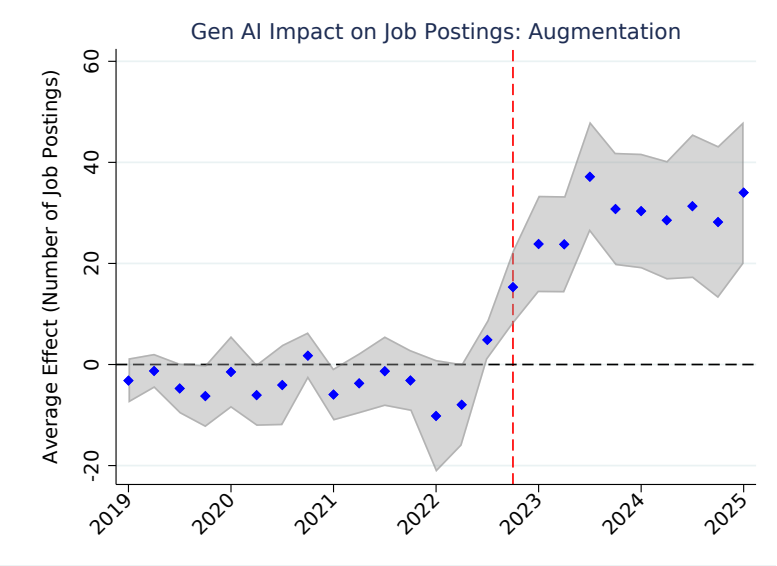
industries. Your task is to analyze a list of skills and classify each into one of four categories:
1. "Gen AI Relevant Skill" (s1) - Software and technical skills directly related to the development, deployment, or application of generative AI technologies.

Examples include: AI ChatBot, AI KIBIT, ANTLR, Apertium, Artificial Intelligence, Automatic Speech Recognition (ASR), Caffe Deep Learning Framework, Chatbot, Computational Linguistics, Computer Vision, Decision Trees, Deep Learning, Google Cloud Machine Learning Platform, Keras, Latent Dirichlet Allocation, Lexical Semantics, Machine Learning, Microsoft Cognitive Toolkit, MLPACK.
2. "Complemented by Gen AI" (s2) - Augmented skills that are “fundamentally human,” but can be enhanced by generative AI tools. Examples include analytical thinking, problem solving, creativity, research, data visualization, strategic planning, predictive analysis, and rapid prototyping.
3. "Substituted by Gen AI" (s3) - Automated skills that will be entirely undertaken by generative AI, with minimal or no human interaction necessary. Examples include image and content generation, data sorting and categorization, forecasting, language translation, simple graphic design, and basic trend spotting.
4. "Irrelevant to Gen AI" (s0) - Limited impact skills that require a human touch, such as complex judgment or nuanced decision-making, that generative AI cannot accomplish. Examples include persuasion and negotiation, motivational leadership, ethical judgment and integrity, compassion, building human relationships, and physical dexterity.
In your classification, consider both the current state of generative AI and emerging trends that might influence the relevance of certain skills in the near future. Provide a classification code (s0, s1, s2, s3) and a brief justification for each classification.

Supplementary Figures and Tables



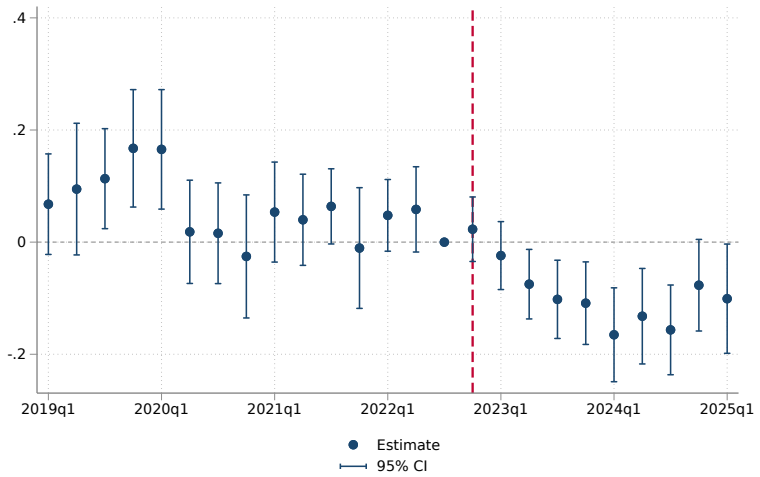
(A) Automation- top quartile of automatable-task share



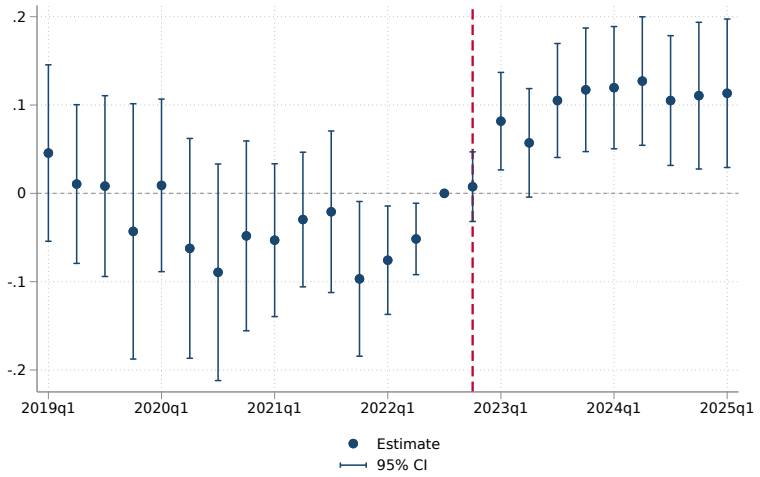
(B) Augmentation- top quartile of automation-exposure diversity

Figure S1. Impact of generative AI on job postings across occupations with high automatable-task share and automation-exposure diversity.

Panel (A) shows the decline in job postings for occupations in the top quartile of the automatable-task share following ChatGPT's release, while Panel (B) illustrates the increase in job postings for occupations in the top quartile of the automation-exposure diversity. Shaded areas represent 95% confidence intervals, and standard errors are calculated using 1,000 bootstrap replications.



(A) Automation- top quartile of automatable-task share



(B) Augmentation- top quartile of automation-exposure diversity

Figure S2. Event Study Estimates from Matched Poisson Difference-in-Differences Models.

Panel (A) plots the estimated dynamic treatment effects on job postings for occupations in the top quartile of the automatable-task share. Panel (B) plots the corresponding estimates for occupations in the top quartile of the automation-exposure diversity. Both panels are based on matched Poisson DiD models using firm-occupation-quarter data. Vertical lines represent 95% confidence intervals, and standard errors are clustered at the firm-occupation level.

Table S1. Distribution of GPT-4o-based skill classifications.

This table presents the distribution of skill-level classification with respect to generative AI, as determined by GPT-4o. Skills are grouped into four categories: S0 (irrelevant to generative AI), S1 (directly relevant to AI development), S2 (complemented by generative AI), and S3 (substitutable by generative AI).

Skill Classification	Skills Count	Skill Share
Skills irrelevant to generative AI (s0)	16,422	48.91%
Generative AI relevant skills (s1)	4,097	12.19%
Skills complemented by generative AI (s2)	12,256	36.45%
Skills substituted by generative AI (s3)	825	2.45%
Total	33,620	100%

Table S2. Impact of ChatGPT's Introduction on Hiring Demand at the Occupation-Level

This table reports robustness estimates on hiring demand from SDID and Poisson DiD models estimated at the occupation-quarter level. The dependent variable is the number of job postings. The treated group consists of occupations in the top quartile of the automatable-task share (Columns 1-2) or automation-exposure diversity (Columns 3-4), while the control group includes occupations in the third and fourth quartiles. SDID estimates (Columns 1 and 3) use a synthetic control group constructed to match pre-treatment trends. Poisson regression estimates (Columns 2 and 4) apply SDID-derived synthetic weights to balance pre-treatment dynamics; observations with zero weights are excluded from the estimation sample, resulting in lower observation counts in the Poisson models. SDID standard errors are bootstrapped with 1,000 replications; Poisson regression standard errors are clustered at the occupation level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Automatable-Task Share		Automation-exposure diversity	
	(1) SDID estimate	(2) DiD estimate (Poisson Regression)	(3) SDID estimate	(4) DiD estimate (Poisson Regression)
Top Quartile x Post-ChatGPT	-1,851***	-0.225***	1,211***	0.197***
	(334.4)	(0.0508)	(426.2)	(0.0407)
Observations	17,175	16,925	17,100	17,100

Table S3. Conventional Difference-in-Differences Analysis of the Impact of ChatGPT's Introduction on Hiring Demand

This table reports conventional DiD estimates of the impact of generative AI on hiring demand, as measured by job postings, for occupations in the top quartile of the automatable-task share (Panel A) and the automation-exposure diversity (Panel B), relative to occupations in the third and fourth quartiles. Columns (1) and (3) present OLS estimates; Columns (2) and (4) present Poisson regression estimates. All specifications include firm-occupation fixed effects and calendar quarter fixed effects. Columns (3) and (4) additionally include industry-by-quarter (NAICS6 × quarter) fixed effects to absorb time-varying sector-level shocks. Standard errors are clustered at the firm-occupation level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. High automatable-task share occupations

	(1)	(2)	(3)	(4)
Top Quartile x Post-ChatGPT	-0.936**	-0.128***	-0.667*	-0.121***
	(-2.55)	(-5.13)	(-1.67)	(-8.64)
Firm-occupation FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Industry×Quarter FE	No	No	Yes	Yes
Observations	7,934,708	7,934,708	7,934,708	7,934,708
Adjusted R^2 /Pseudo R^2	0.558	0.830	0.558	0.842

Panel B. High automation-exposure diversity occupations

	(1)	(2)	(3)	(4)
Top Quartile x Post-ChatGPT	2.515***	0.184***	2.244***	0.156***
	(7.33)	(8.39)	(6.39)	(9.04)
Firm-occupation FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Industry×Quarter FE	No	No	Yes	Yes
Observations	7,940,935	7,940,935	7,940,935	7,940,935
Adjusted R^2 /Pseudo R^2	0.609	0.856	0.610	0.866

Table S4. Matched Difference-in-Differences Analysis on the Impact of ChatGPT's Introduction on Hiring Demand

This table reports DiD estimates on hiring demand, as measured by job postings, using matched samples of firm-occupation-quarter observations. The treated group includes occupations in the top quartile of the automatable-task share (Panel A) or the automation-exposure diversity (Panel B), matched to control occupations in the third and fourth quartiles. Matching is performed exactly on industry (NAICS-2), firm size (employee count group), and Job Zone, followed by propensity score matching on pre-treatment job postings (nearest-neighbor with replacement, two matches per treated unit). Columns (1) and (3) report OLS estimates; Columns (2) and (4) report Poisson regression estimates. All specifications include firm-occupation fixed effects and calendar quarter fixed effects. Columns (3) and (4) additionally include industry-by-quarter (NAICS6 × quarter) fixed effects. Standard errors are clustered at the firm-occupation level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A. High automatable-task share occupations

	(1)	(2)	(3)	(4)
Top Quartile x Post-ChatGPT	-1.855***	-0.194***	-1.793***	-0.170***
	(-5.79)	(-8.79)	(-5.63)	(-9.45)
Firm-occupation FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Industry×Quarter FE	No	No	Yes	Yes
Observations	4,973,448	4,973,448	4,973,448	4,973,448
Adjusted R^2 /Pseudo R^2	0.593	0.801	0.594	0.815

Panel B. High automation-exposure diversity occupations

	(1)	(2)	(3)	(4)
Top Quartile x Post-ChatGPT	2.399***	0.130***	2.145***	0.119***
	(5.79)	(5.74)	(4.95)	(6.07)
Firm-occupation FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Industry×Quarter FE	No	No	Yes	Yes
Observations	3,103,849	3,103,849	3,103,849	3,103,849
Adjusted R^2 /Pseudo R^2	0.662	0.870	0.663	0.881

Table S5. Restricted Firm Sample Analysis of the Impact of ChatGPT's Introduction on Hiring Demand

This table reports the effects of ChatGPT's introduction on hiring demand, as proxied by job postings, for occupations in the top quartile of the automatable-task share (Columns 1-2) and the automation-exposure diversity (Columns 3-4), relative to those in the third and fourth quartiles. The sample is restricted to firms that posted both treated and control occupations during the study period. SDID estimates (Columns 1 and 3) use a synthetic control group constructed to match pre-treatment trends. Poisson regression estimates (Columns 2 and 4) apply SDID-derived synthetic weights to balance pre-treatment dynamics; observations with zero weights are dropped, resulting in a smaller estimation sample. SDID standard errors are bootstrapped using 1,000 replications, and Poisson regression standard errors are clustered at the firm level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Automatable-Task Share		Automation-exposure diversity	
	(1) SDID estimate on Job Postings	(2) DiD estimate on Job Postings (Poisson Regression)	(3) SDID estimate on Job Postings	(4) DiD estimate on Job Postings (Poisson Regression)
Top Quartile x Post-ChatGPT	-27.46***	-0.127**	28.74***	0.205***
	(5.634)	(0.0542)	(5.777)	(0.0279)
Observations	503,225	425,225	563,825	563,825

Table S6. Impact of ChatGPT's Introduction on Hiring Demand Based on Alternative Definitions of the Automatable-Task Share

This table reports the effects of ChatGPT's introduction on hiring demand, as proxied by job postings, for occupations in the top quartile of the automatable-task share under two alternative constructions, relative to those in the third and fourth quartiles. Columns (1)-(2) define the automatable-task share as the importance-weighted share of tasks classified E1+E2; Columns (3)-(4) define it as E1+E2+E3. SDID estimates (Columns 1 and 3) use a synthetic control group constructed to match pre-treatment trends. Poisson regression estimates (Columns 2 and 4) apply SDID-derived synthetic weights to balance pre-treatment dynamics; observations with zero weights are dropped, resulting in a smaller estimation sample. SDID standard errors are bootstrapped using 1,000 replications, and Poisson regression standard errors are clustered at the firm-quartile exposure level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	E1+E2		E1+E2+E3	
	(1) SDID estimate on Job Postings	(2) DiD estimate on Job Postings (Poisson Regression)	(3) SDID estimate on Job Postings	(4) DiD estimate on Job Postings (Poisson Regression)
Top Quartile x Post-ChatGPT	-31.00***	-0.135**	-32.03***	-0.147***
	(10.04)	(0.0543)	(9.663)	(0.0548)
Observations	526,825	441,600	523,825	441,725

Table S7. Impact of ChatGPT's Introduction on Hiring Demand Based on the Anthropic Economic Index (AEI).

This table reports the effects of ChatGPT's introduction on hiring demand, as proxied by job postings, for occupations in the top quartile of AEI-based automation (Columns 1-2) and AEI-based augmentation (Columns 3-4), relative to those in the third and fourth quartiles. AEI task records are mapped to *O*NET* tasks; tasks with AEI mass ≥ 0.75 (share of valid, non-filtered interactions) are retained and aggregated to occupations using *O*NET* Importance weights. SDID estimates (Columns 1 and 3) use a synthetic control group constructed to match pre-treatment trends. Poisson regression estimates (Columns 2 and 4) apply SDID-derived synthetic weights to balance pre-treatment dynamics; observations with zero weights are dropped. SDID standard errors are bootstrapped using 1,000 replications; Poisson standard errors are clustered at the firm-quartile exposure level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	AEI Automation		AEI Augmentation	
	(1) SDID estimate on Job Postings	(2) DiD estimate on Job Postings (Poisson Regression)	(3) SDID estimate on Job Postings	(4) DiD estimate on Job Postings (Poisson Regression)
Top Quartile x Post-ChatGPT	-2.948	-0.0603	12.68***	0.137***
	(2.776)	(0.0419)	(3.241)	(0.0392)
Observations	564,525	441,450	566,300	436,800