AI and Jobs: Has the Inflection Point Arrived? Evidence from an Online Labor Platform

Dandan Qiao

School of Computing, National University of Singapore

Email: qiaodd@nus.edu.sg

Huaxia Rui

Simon Business School, University of Rochester

Email: huaxia.rui@simon.rochester.edu

Qian Xiong

Institute of Project Management and Construction Technology, Tsinghua University

Email: xq20@mails.tsinghua.edu.cn

Abstract

Artificial intelligence (AI) refers to the ability of machines or software to mimic or even surpass human intelligence in a given cognitive task. While humans learn by both induction and deduction, the success of current AI is rooted in induction, relying on its ability to detect statistical regularities in task input — an ability learnt from a vast amount of training data using enormous computation resources. We examine the performance of such a statistical AI in a human task through the lens of four factors, including task learnability, statistical resource, computation resource, and learning techniques, and then propose a three-phase visual framework to understand the evolving relation between AI and jobs. Based on this conceptual framework, we develop a simple economic model of competition to show the existence of an inflection point for each occupation. Before AI performance crosses the inflection point, human workers always benefit from an improvement in AI performance, but after the inflection point, human workers become worse off whenever such an improvement occurs. To offer empirical evidence, we first argue that AI performance has passed the inflection point for the occupation of translation but not for the occupation of web development. We then study how the launch of ChatGPT, which led to significant improvement of AI performance on many tasks, has affected workers in these two occupations on a large online labor platform. Consistent with the inflection point conjecture, we find that translators are negatively affected by the shock both in terms of the number of accepted jobs and the earnings from those jobs, while web developers are positively affected by the very same shock. Given the potentially large disruption of AI on employment, more studies on more occupations using data from different platforms are urgently needed.

Keywords: artificial intelligence, online labor market, jobs, tasks, ChatGPT

1 Introduction

Probably the two most important aspects of human intelligence are induction and deduction. These were recognized as early as by Aristotle in his *Posterior Analytics* where he contrasted the "syllogistic and inductive." To replicate or surpass human intelligence using machines or software, early research in artificial intelligence (AI) focused on teaching machines to do deduction (e.g., automated reasoning and knowledge representation) which only resulted in limited progress. In contrast, recent decades have witnessed enormous success in training machines to do induction, popularly known as machine learning. As significant and dominant as this progress is, it is important to recognize that the success of current AI technologies is largely based on the detection of statistical regularities in data, rather than from step-by-step rigorous deduction using logic. If human intelligence is indeed a mixture of System 1 (featuring fast thinking) and System 2 (featuring slow thinking), as suggested by Kahneman (2013), then current AI technologies can be viewed as fast thinking pushed to the limit through the use of enormous amounts of data and intensive computation. To distinguish the pursuit of these two forms of intelligence, we refer to them as statistical artificial intelligence (Statistical AI) and causal artificial intelligence (Causal AI), respectively. Such a distinction is important for the current paper because the ingredients and underlying logic of our proposed conceptual framework are based on and only apply to Statistical AI. For ease of exposition, we refer to Statistical AI simply as AI from now on.

The objective of the present paper is twofold. First, we take a closer look at factors that determine successful applications of AI in the production of goods and services, which naturally leads to a conjecture and related hypothsis regarding AI's impacts on various occupations. A closely related question of how automation technology affects employment and wages has been extensively studied in macroeconomics and labor economics. In this fruitful literature, technology itself is treated as a black box, entering an economy's production function as a factor alongside human labor in an aggregated manner. This macroscopic approach is essentially technology-agnostic and focuses on the long-term impact of any automation technology. Because our focus is on a particular technology, namely AI, we dive deeper into the crucial factors that determine the ability of the technology to complete various tasks. We believe this is warranted because AI is a general-purpose technology that may revolutionize how we live and work in the next few decades. Moreover, the power of AI depends crucially on the interplay among several factors that are unique to this technology.

More specifically, we suggest a conceptual framework where any cognitive task performed by humans in the production of goods or services is represented by a function f that maps certain task inputs (e.g., paragraphs in one language, driving conditions) X to some desirable task output $Y = f(X)$ such as the correct translation in another language or adequate vehicle controls. AI approximates human-level intelligence by learning from a large amount of labeled data $\{(x_i, y_i)\}$ which are often collected from successful human processes corresponding to various task inputs. Although tasks differ from each other in a multitude of dimensions, central to AI's ability to learn a task is the task's learnability which is an inherent feature of the task. The notion of task learnability reflects two conditions for induction to be successful. First, because the success of induction relies on generalization from observed instances to unknown instances, it is essential that statistical regularity exists in and can be extracted from the data. While the existence of statistical regularity should be satisfied for data naturally generated from the production of goods and services, the amount of data needed to extract the statistical regularity may vary greatly depending on the underlying task. Second, it must be computationally feasible for a learning algorithm to approximate f by detecting and exploiting the underlying statistical regularity. Just because f is polynomial time does not mean the learning algorithm itself is polynomial time. Indeed, successful learning is not guaranteed even with an arbitrary number of labeled instances at our disposal, as is evidenced by the practical impossibility of known-plaintext attack against key public infrastructure — the backbone of trust on the Internet. Hence, we break down the learnability of a task f into its statistical and computational aspects, denoted by S_f and C_f, respectively, and represent task f as the point (S_f, C_f) on a task plane.

Adding the intelligence level as the height dimension to the task plane, we obtain the task intelligence space where the intelligence level can be absolute or relative (e.g., a percentage of the best human performance for the task). We imagine a surface in the task intelligence space that represents the current intelligence level achieved by AI for all cognitive tasks. We refer to such a surface as the current intelligence surface (CIS). Whenever more training instances become available for a particular task, the intelligence surface corresponding to that task and other similar tasks (via transfer learning) bumps up, especially if the task is high in S_f and low in C_f. Similarly, whenever there is an improvement in computation speed (e.g., graphics processing unit, or GPU), those tasks high in C_f and low in S_f will experience a particularly strong boost in intelligence level. Innovations in learning algorithms for certain tasks can lift the region of the intelligence surface corresponding to those tasks, by better exploiting the data, searching more efficiently in the hypothesis class, or discovering a better hypothesis space. Recent examples of such innovations include new training techniques such as dropout and new hypothesis classes such as convolutional neural networks (CNN) for vision tasks and transformer architecture for natural language processing (NLP) tasks.

Because a job typically involves multiple tasks, we visualize a job category or occupation as a set of points (henceforth task set) on the task plane. If the CIS is above the minimal intelligence surface required for satisfactory completion of all tasks in the task set of an occupation, AI can substitute human labor in such an occupation. If the CIS is above the minimal intelligence surface at some but not all points in the task set, AI can complement human workers by making them more productive while data generated from human workers can improve AI performance. If the CIS is below the minimal intelligence surface for all points in the task set, AI has no effect on such an occupation. Correspondingly, we may categorize the relation between AI and an occupation into three phases: the decoupled phase, the honeymoon phase, and the substitution phase. Whenever there is an improvement in the three variable factors (i.e., data availability, computing speed, and learning technologies), the AI-jobs relation may transition from the decoupled phase to the honeymoon phase, or from the honeymoon phase to the substitution phase. Moreover, once the relation enters the honeymoon phase, more labeled data can be obtained through the use of human labor and AI being deployed side by side, which in turn should speed up the improvement in AI, thereby accelerating further transition from the honeymoon phase to the substitution phase. Because technology diffusion takes time, different organizations may have different CISs. Even for the same occupation, its relation with AI may be in the substitution phase for some organizations but in the honeymoon phase for others. Therefore, to apply the framework in practice, the trichotomy of the three phases should be interpreted probabilistically. Analyzing a simple economic model based on this framework, we show the existence of an inflection point for each occupation after which human workers become worse off whenever AI improves, in stark contrast to the period before the inflection point when human workers benefit from improvement in AI.

Our second objective is to test the above inflection point conjecture using empirical data. For that, we conducted an empirical study using the launch of ChatGPT on November 30, 2022, as an exogenous shock that raised certain regions of the CIS. Our data comes from a large online freelancing platform, and we focus on two job categories: translation and web development. We hypothesize that the occupation of translation has passed the inflection point, especially after the launch of ChatGPT, while the occupation of web development has not, even after the launch of ChatGPT. For translation, we note that the transformer architecture, which is at the core of GPT, was initially proposed to tackle the challenge of machine translation. In contrast, web development involves high-level designs and complex interactions among different components, making it more challenging for AI to completely surpass the minimal intelligence levels of tasks required for web development. On the other hand, AI tools like ChatGPT do make the job of programming more efficient by assisting programmers with debugging, code snippets, and so on. Consequently, we believe the launch of ChatGPT did shock the area of the CIS corresponding to web development. As a control group, we use the occupation of construction design because jobs in this category are currently completed by humans using specialized software with very limited inputs from AI, most likely because of insufficient data for training AI.

We match workers in a treated occupation (i.e., translation, web development) with workers in the control occupation (i.e., the construction design) and conduct a difference-in-differences (DID) analysis at the worker-month level. Our first dependent variable is transaction volume measured either as a worker's accepted number of focal jobs each month or the ratio between those focal jobs and the total number of jobs accepted by the worker each month. Our second dependent variable is total earnings from those focal jobs. Consistent with the inflection point conjecture, we find that for translators, their transaction volumes dropped after the launch of ChatGPT, and they earned less from translation jobs. In contrast, web developers experienced an increase in their transaction volume and earnings after the launch of ChatGPT.

The remainder of the paper is organized as follows. Section 2 reviews several streams of literature related to our paper. In Section 3, we develop our conceptual framework. In Section 4, we elaborate on the empirical context and report findings from the empirical analyses. In Section 5, we extend the main analyses by examining three more job categories on the platform: writing, machine learning, and physical sciences, to further test our theory. Finally, we conclude the paper in Section 6 by summarizing the contributions and discussing the limitations and potential future research directions.

2 Research Background

2.1 Impact of Automation Technology on Labor Market

In the past decades, automation technology has seen tremendous development, raising concern in relation to "technological unemployment". To a large extent, automation technology eliminates the demand for labor undertaking repeated and manual work. Such a substitution has shifted the labor demand towards skilled and highly educated ones (Autor et al. 1998).

However, at the same time, researchers have also acknowledged automation technology as an effective tool to augment human ability, enhancing their competence in the labor market (Autor et al. 2003). Some studies further demonstrated that these technologies have the potential to create new industries and job opportunities for human labor (Acemoglu and Restrepo 2019). These mixed effects (i.e., substitution and complementarity effects) give rise to an important research branch exploring the relation between automation technology and labor.

Economists have engaged in extensive theoretical deliberation to understand how automation technology might impact human labor. Some research utilizes economic models to describe the elasticity of substitution among different production factors, such as IT, labor, and capital (Dewan and Min 1997, Zhang et al. 2015). Other research has extensively explored the role of technology in working processes. Notably, Autor et al. (2003) introduced the perspective of job task composition to explain how computer technology alters tasks within an occupation and subsequently affects the demand for human skills. Specifically, routine tasks, governed by explicit rules, are readily automated, whereas nonroutine tasks, lacking defined rules, primarily experience a complementarity effect with automation technologies. This "Routine-biased Technological Change" perspective is widely acknowledged for understanding how technological change impacts various types of human labor.

Empirical evidence on the impact of automation technology on labor markets presents divergent findings. At the aggregate level, while some found a net substitution effect (Chwelos et al. 2010), some found evidence for a net complementarity effect (Bresnahan et al. 2002). At the micro level, however, the impact often depends on different types of employers or workers (Lu et al. 2018, Zhang et al. 2023). For instance, Lu et al. (2018) showed in the context of health IT adoption, that licensed nurse staffing levels increased in low-end nursing homes but decreased in high-end nursing homes. Zhang et al. (2023) proved that while highly educated labor received a complementarity effect and lowly educated labor received a substitution effect, the net effects on averagely educated labor depended on task routineness.

Recently, academia's attention has shifted towards AI, due to its increasing role in our society. The first stream of research expands upon prior theoretical frameworks to enhance the understanding of AI-labor relation. Employing a task-based approach, Acemoglu and Restrepo (2019) posited that automation, specifically AI and robotics, extensively displaces human labor. Nonetheless, they also emphasize the presence of countervailing aspects with the potential to mitigate this displacement effect. Acknowledging AI's premier capability in prediction, Agrawal et al. (2019) delineated jobs into prediction and decision tasks, suggesting that AI's impact on various job categories could be ambiguous. These studies build their frameworks by emphasizing that AI only automates certain categories of task, but fell short of revealing the key factors determining AI's successful application to these specific tasks. This gap could stem from a disregard for AI's developmental characteristics, critical in shaping its relation with labor, which our research endeavors to resolve.

The second strand of research provides empirical evidence of AI's dual effects on the labor market, aligning with findings observed in broader automation technology studies. For instance, Lysyakov and Viswanathan (2022) revealed that lower-tier designers tend to exit the online market when facing the threat of image-generating AI, while high-tier designers could become more engaged. Xue et al. (2022) demonstrated that increasing AI applications positively impact the employment of non-academically trained workers in firms, yet adversely affect academically trained employees, which collectively indicates a net positive effect on overall employment. However, these studies primarily rely on data from single occupations or macro-level analysis, potentially overlooking AI's heterogeneous impact across various labor markets. This has also been recognized as a significant barrier in comprehending the contextual impact of AI on the workforce (Agrawal et al. 2019). Our research, capitalizing on the advent of recent LLM technology, tackles this issue by examining a unique worker-level dataset, which covers a broad spectrum of job categories.

2.2 Online Labor Market

The online labor market (OLM) has grown tremendously in the past decades. The OLM has shifted the traditional labor market onto online platforms, introducing new avenues for labor transactions in the digital economy. By joining an OLM, workers can access job opportunities beyond national boundaries, actively participating in the global labor market instead of being confined to local demand (Kanat et al. 2018). The emergence of this market also benefits employers by enabling platform-mediated transactions and communication, thereby reducing transaction costs (Horton 2010). By 2021, more than 160 million user accounts have been registered as online freelancers [1].

Existing literature on OLM can be categorized into three streams, corresponding to the focus on workers, employers, and the platform. From the perspective of labor supply, OLM is an alternative marketplace for employment and serves as an influential and effective offset for offline unemployment (Huang et al. 2020). Researchers also focus on workers' well-being, highlighting the significant roles of reputation and skills in determining their market value (Lin et al. 2018, Kokkodis 2023). From the perspective of labor demand, existing literature mainly tries to answer how an employer can optimize the hiring decision. A key factor is the employer's reputation, aiding in attracting superior talent and streamlining transaction and negotiation processes (Benson et al. 2020). Studies also emphasize hiring as a dynamic learning process, stressing the value of insights gained from prior hiring experiences (Leung 2018). From the platform's standpoint, academic research primarily concentrates on fostering effective communication between online employers and workers as well as optimizing operations, such as strategies for platform incentives and bid auctions (Barach et al. 2020, Hong et al. 2016).

OLM's basis on AI-exposed digital platforms and the inherent flexibility of online workers magnify the immediacy and extensive impact of automation technology (Horton 2010). This spurs a recent wave of literature dedicated to algorithm-based features to facilitate employee-

[1] Oxford Internet Institute: https://ilabour.oii.ox.ac.uk/how-many-online-workers/

employer matching from the perspective of platform operations (Horton 2017, Kokkodis and Ipeirotis 2023). For instance, Horton (2017) conducted a field experiment and demonstrated that algorithmic recommendation could significantly help employers fill their online technical job vacancies. Kokkodis and Ipeirotis (2023) considered job-application characteristics to further improve the recommendation system for OLMs. However, in recent years, there has been a notable enhancement in AI's proficiency to emulate and comprehend human language and emotions (Yu et al. 2023, Wies et al. 2023). This empowered AI as an active participant in the labor market alongside workers, instead of a mere operational tool. Our research is hence motivated to investigate how AI impacts workers in the OLMs.

2.3 Large Language Models (LLMs)

Large Language Models (LLMs) have emerged as a revolutionary advancement in the realm of AI. The development of LLMs aims to address limitations in existing machine learning (ML) systems, which rely on supervised learning for language understanding (Radford et al. 2019). These conventional ML systems typically function as supervised learners, which are trained from limited-domain datasets and are sensitive to data distributions, resulting in their lack of generalization. LLMs have freed themselves from reliance on explicit supervision and are instead pretrained on extensive general-purpose internet data to achieve the goal of maximally mimicking human language. In this pretraining process, LLMs naturally assimilate all relevant linguistic information and knowledge for language generation, which endows LLMs with innate abilities to process various downstream applications (Brown et al. 2020). For instance, LLMs are frequently utilized for the efficient completion of tasks like translation and writing by analyzing the given prompts, as evidenced in prior work that highlights their use in assisting with ad copy creation (Chen and Chan 2023). This is known as "in-context learning" (Wies et al. 2023), which means that LLMs can adapt to diverse tasks without altering their internal structure, merely by integrating specific instructions or examples within their input.

Studies have attempted to both practically and theoretically explain the mechanisms behind the "in-context learnability" of LLMs, as discussed in works by Brown et al. (2020) and Radford et al. (2019). Despite being initially configured to maximize the probability of predicting unlabeled internet texts during pretraining, LLMs inherently acquire a wide array of abilities for language understanding and relevant task execution. Once these competencies are acquired and embedded through pretraining, "in-context learning" in LLMs primarily involves recognizing and applying these capabilities in response to specific instructional inputs for varied tasks (Wies et al. 2023). This method closely mirrors the human approach to task processing, where understanding and action are derived directly from textual instructions.

The emergent abilities endowed by the pretraining process allow LLMs to contribute to various labor sectors. A notable instance is the launch of ChatGPT, namely Chat Generative Pre-trained Transformer, which appears to hold substantial sway in the labor markets. ChatGPT was the first to bring the application of LLMs to the general public, and it has rapidly become a valuable tool for individuals and organizations. Since its launch, ChatGPT has reportedly amassed around 100 million active users monthly, setting a new record as the quickest-growing consumer app ever. Careers from different domains have been exposed to this popular AI tool (Eloundou et al. 2023), alarming people to rethink the "technology displacement" issue.

On the one hand, LLMs have the potential to act similarly to human workers by interpreting and executing tasks based solely on text-based instructions. As cost-effective and high-quality labor alternatives, LLMs might pose a significant challenge to the role of and even the necessity for human labor in certain markets (Eloundou et al. 2023). On the other hand, the evolution of LLMs is leaning towards reducing barriers to entry into various labor sectors by enhancing AI's comprehension capabilities (Wies et al. 2023), potentially benefiting employees across diverse skill levels. While numerous debates and discussions have taken place, there remains a lack of empirical investigation into the impact of ChatGPT on the labor market.

To the best of our knowledge, the closest work to ours is a concurrent working paper by Liu et al. (2023), which investigates how the launch of ChatGPT affected transaction volume on an online labor platform. Their main finding is a significant decrease in transaction volume for gigs and freelancers directly exposed to ChatGPT. In contrast, our study reveals a more complex relation between AI and jobs, both theoretically and empirically. In particular, we propose the inflection point conjecture based on our conceptual framework and economic modeling that leads to different implications of ChatGPT's launch on different job categories. It should be noted that our findings do not contradict the main finding of Liu et al. (2023), because their treated job category consists of writing and programming which, according to our analyses, should have experienced opposite effects with the launch of ChatGPT. It's likely that the platform in their study differs from the platform in our study, which we believe makes the two studies complementary to each other for the robustness of empirical research in our field.

3 Hypotheses Development

We develop our hypotheses in three steps. First, to understand AI performance, we suggest a four-factor framework — task learnability, statistical resource, computation resource, and learning techniques — where task learnability is an inherent property of each task consisting of the task's statistical complexity and computational complexity. Representing each task as a point on the task plane spanned by these two dimensions, we may visualize current AI performance as a surface above the task plane. Second, we consider an occupation as a task set on the task plane and compare the relative position of the CIS with the minimal intelligence surface. This naturally leads to a three-phase relation between AI and jobs, allowing us to propose a conjecture that categorizes the impact of AI progress on an occupation based on whether AI has crossed the inflection point of that occupation. Finally, we develop two testable hypotheses based on this inflection point conjecture.

3.1 AI and Tasks Through the Lens of Four Factors

We conceptualize a cognitive task or task for short, as the smallest unit in the production of certain goods or services. A task can be represented as a function f mapping a certain task input X to some desirable output $Y = f(X)$, or more generally, to a probability distribution over possible outputs. For humans, such a mapping is implemented in biological neural networks and is acquired through, for most tasks, years of observing, thinking, and practicing. Moreover, the hypothesis class $\mathcal{H}$ and the learning algorithm $\mathcal{A}$ humans use for learning are inherited from tens or even hundreds of millions of years of evolution.

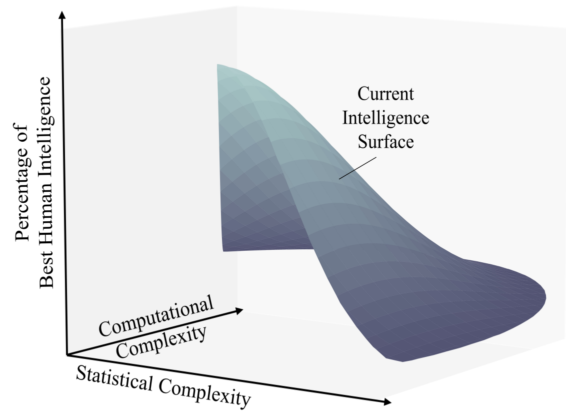
For an AI to learn this mapping f, we need to provide it with both statistical and computational resources. The volume of statistical resources required is clearly task dependent. For example, learning whether to switch lanes should require more labeled instances than learning whether to stop at a traffic light. We conceptualize a measure of statistical complexity for a task f and denote it by S_f. At an intuitive level, we may think of the statistical complexity of a task as the "inverse" of its statistical regularity which pertains to the consistent patterns or traits that a task exhibits when observed repeatedly. A task that demonstrates strong regularities in the training data tends to be more readily learned by a learning algorithm, which can then exploit these regularities to make predictions on unseen data. Literature in economics often characterizes tasks by their level of routineness which can be thought of as a simplified version of statistical regularity because routine tasks are considered so predictable that they can be implemented by manually crafted rules without using much data. More precisely, we may define S_f as the smallest number of labeled instances needed for the AI to approximate f sufficiently well. Recall that in Probably Approximately Correct (PAC) learning theory (Shalev-Shwartz and Ben-David 2014), given $\epsilon, \delta \in (0, 1)$ and a learning algorithm, the sample complexity of learning a hypothesis class is the smallest training sample size that a learning algorithm needs to produce a predictor such that the generalization error is less than ϵ with a probability of at least $1 - \delta$. In our context, we may define S_f as the sample complexity associated with $\mathcal{H}$ and $\mathcal{A}$ for some fixed performance threshold (ϵ, δ). In practice, the minimal number of labeled instances required for an AI to learn f is likely much larger than S_f because we know neither $\mathcal{H}$ nor $\mathcal{A}$. Two main sources of AI innovation are better hypothesis classes and more efficient and effective learning algorithms.

Like statistical resources, the amount of computation resources required is also task dependent. For example, tasks involving images and videos are typically more computationally intensive than tasks involving only numerical features. The low statistical complexity of a task does not imply the task can be easily learnt. In fact, for some polynomial-time tasks, even with an arbitrary number of labeled instances at our disposal, there may not exist any polynomial-time learning algorithm to acquire f, making learning infeasible. Hence, we conceptualize a measure of computational complexity, denoted by C_f, for the *learning* of task f without knowing $\mathcal{H}$ and $\mathcal{A}$. The larger C_f is, the harder it is for AI to learn f.

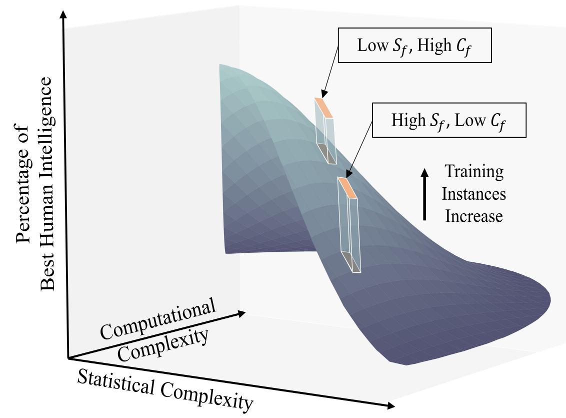
Both the statistical complexity and the computational complexity of a task are inherent properties of the task, and we refer to them collectively as the task's learnability. If we imagine a task plane where each task is represented as a point[2] with coordinates (S_f, C_f), then at any moment in time, the levels of intelligence achieved by AI for all tasks form a surface above the task plane, in a task intelligence space. Figure 1(a) illustrates such a surface which we refer to as the current intelligence surface (CIS). If all tasks share the same number of labeled data, then the intelligence surface should be a smooth surface, assuming equal access to the same computing hardware.

Over time, some part of the surface is raised upwards, reflecting the performance improvement of AI on certain tasks. We distinguish three types of shifts. First, an increase in labeled data can lead to an increase in AI performance due to the reduction of estimation errors caused by finite samples. AI performance on tasks with high statistical complexity but moderate computational complexity can gain the most from increased data availability.

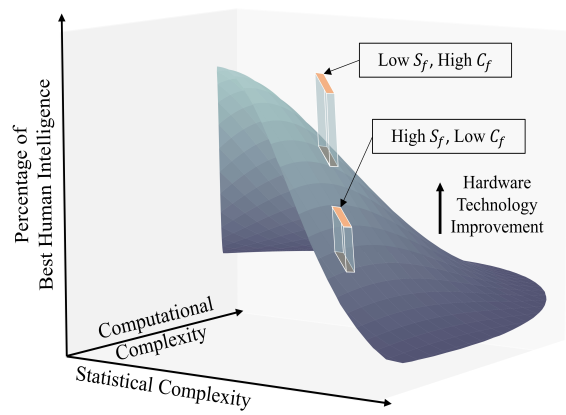
[2] Without loss of generality, we may assume a one-to-one mapping between tasks and points on the task plane for ease of visualization.



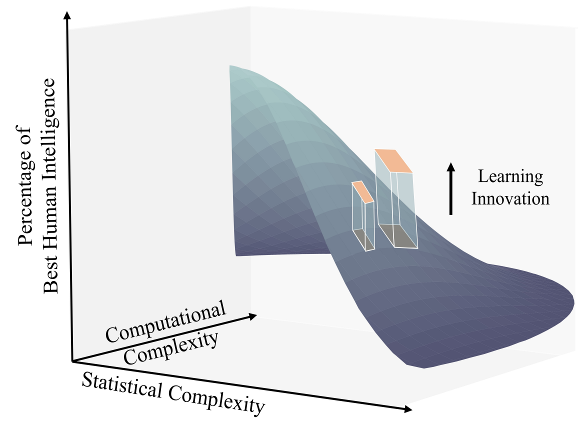
(a) Task Intelligence Space and Current Intelligence Surface (CIS)



(b) Shocking the CIS: Role of Data



(c) Shocking the CIS: Role of CPU/GPU/TPU



(d) Shocking the CIS: Role of Learning Innovations

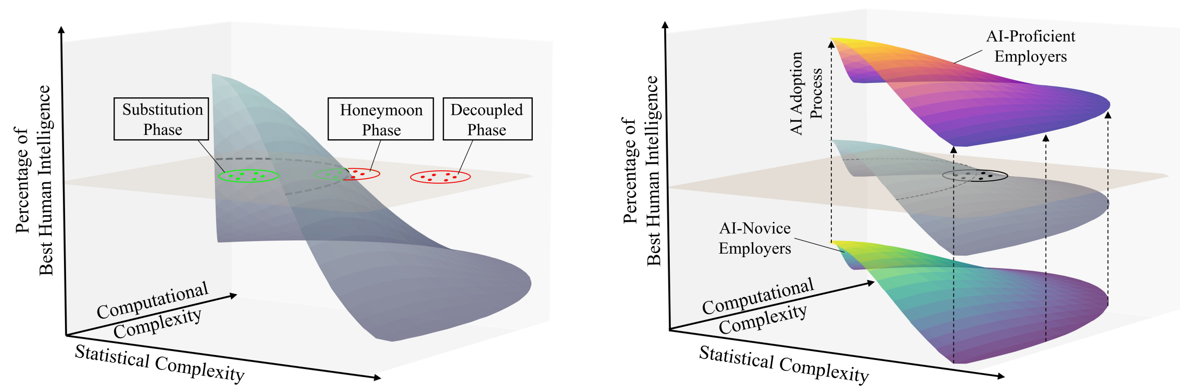
Figure 1: Current Intelligence Surface and the Four Determinants of AI Performance

Moreover, there may be a spillover effect from increased data availability for one task to other similar tasks, thanks to transfer learning. Figure 1(b) illustrates this. One recent example of such a scenario is the ImageNet project launched in 2009 which contains more than 14 million annotated images of over 20,000 categories. The project significantly contributed to the rapid advances of AI performance on numerous computer vision tasks. Second, progress in computing technology such as GPU can speed up the search for a better approximation of the task mapping f, leading to an improvement in AI performance. Tasks with high computational complexity but moderate statistical complexity should benefit the most from increased computation speed. Figure 1(c) illustrates this. Finally, learning innovation, in terms of new learning algorithms or better hypothesis classes, can raise part of the CIS, even in the absence of any increase in data availability or computation speed. For example, regularization techniques can help an Empirical Risk Minimization (ERM) learner find a better predictor than it would otherwise find without regularization. Different tasks often require different regularization techniques although some regularization methods are more general than others. Similarly, using a better hypothesis class for certain types of tasks can significantly improve the performance of AI on those tasks. Recent examples include various innovations in neural network architectures (e.g., AlexNet in 2012 and Transformer in 2017) that have led to significant performance improvements in many tasks. Because learning innovations can be either task-specific or general-purpose, the affected region of the intelligence surface may be small or large, as is illustrated in Figure 1(d).

3.2 A Three-phase Relation Between AI and Jobs

To understand the implication of AI on jobs, we first conceptualize an occupation as a task set, represented as a set of points on the task plane. For example, the job of translating a document from one language to another may involve the task of literal translation and the task of localization. Second, for a task to be successfully completed by an AI, the AI performance needs to surpass a certain threshold (e.g., 90% of the best human worker). Different tasks may require different thresholds. Hence, we conceptualize a minimal intelligence surface to represent this threshold for various tasks on the task plane. We analyze the relation between AI and an occupation by comparing the relative position of the CIS and the minimal intelligence surface in the region corresponding to the task set of the job. Figure 2(a) illustrates three scenarios where, for convenience, we plot the minimal intelligence surface as a flat surface.

On one extreme, visualized as the task set on the left, the CIS is completely above the minimal intelligence surface on the task set corresponding to an occupation. We argue that such an occupation is at risk of becoming obsolete because AI can perform as well as humans but at a much smaller marginal production cost. We refer to this scenario as the



(a) The Three Phases of AI-Jobs Relation (b) AI Adoption Takes Time

Figure 2: The Evolving Relation Between AI and Jobs

substitution phase. Once AI's relation with an occupation enters the substitution phase, it is more efficient for employers to employ AI than humans for such jobs. At the other end of the spectrum is the scenario where the CIS is completely below the minimal intelligence surface in the region corresponding to the task set of an occupation. In such a scenario, AI is irrelevant because not a single task of those jobs can yet be satisfactorily completed by AI. We refer to this scenario as the decoupled phase because human workers are not engaging with AI while doing these jobs.

Between these two extremes, we have the most interesting scenario where the CIS is above the minimal intelligence surface on some tasks of an occupation but is below the minimal intelligence surface on other tasks of the occupation. In other words, these jobs still have to be done by human workers but AI can help by satisfactorily completing some of the tasks. We refer to this scenario as the honeymoon phase because human workers enjoy enhanced productivity by engaging with AI while AI benefits from increased data availability by working side-by-side with humans.

Because the minimal intelligence surface is static, as the CIS shifts upwards over time, AI's relation with an occupation naturally moves from the decoupled phase to the honeymoon phase, and may eventually transition into the substitution phase for some occupations. The transition speed and the relative duration in each phase likely vary depending on the occupation and the progress of AI. However, there is likely to be an acceleration once the relation enters the honeymoon phase because more labeled data can be generated from humans working closely with AI, which in turn should speed up the improvement in AI.

Our discussions so far are based on only one intelligence surface, representing the CIS at the societal level. In reality, different organizations or different regions may have different CISs because technology diffusion takes time. Even for the same occupation, its relation with AI may be in different phases for different organizations, as is illustrated in Figure 2(b). Hence, the trichotomy of these three phases should be interpreted probabilistically. In other words, each occupation's relation with AI can be characterized by a probability distribution over the three phases, reflecting the proportion of organizations for which the relation is in each phase.

To formalize the above analyses, consider a Cournot competition model with n workers each providing the same service with the same marginal cost of producing one unit of service. Let the marginal cost be $(1-a)c$ where $c > 0$ and $a \in [0, 1]$. We interpret a as the percentage of tasks that can be successfully completed by AI during the production of the service. So c represents a worker's marginal cost without using any AI assistance. Market demand for the service is determined by $p = S(a) - \sum_i q_i$ where p is the price, q_i is the quantity of services provided by worker i, and $S(a)$ represents the market potential. $S(a)$ is decreasing in a. To see this, consider the heterogeneity among potential employers in terms of their CISs. For potential employers who are more AI literate, their CISs are above the minimal intelligence surface for the focal job, hence will substitute AI for labor. As AI improves, i.e., an increase in a, more potential employers fall into that category, thereby reducing the market potential. Moreover, $S(a)$ is likely concave because technology adoption often accelerates as the technology matures. There are several possible mechanisms. First, as AI performance increases, more employers will use it which creates more word-of-mouth recommendations, hence more adoptions. Second, there is a positive externality from more employers using AI due to the dissemination of know-how and best practices. Third, innovative businesses will develop specialized software to facilitate the use of AI to aid specific occupations, as AI becomes increasingly powerful for that type of job. We impose the technical assumptions of $|S'(0)| < c$ and $|S'(1)| > c$ to avoid non-interesting cases.

Each worker maximizes profit $\pi_i = pq_i - (1-a)cq_i = S(a)q_i - q_i^2 - \sum_{j\neq i} q_j q_i - (1-a)cq_i$, which yields the first-order condition

$$\frac{\partial \pi_i}{\partial q_i} = \big(S(a) - (1-a)c\big) - 2q_i - \sum_{j\neq i} q_j = 0 \iff q_i^* = \frac{S(a) - (1-a)c - \sum_{j\neq i} q_j}{2}$$

We can easily solve for the Nash equilibrium as:

$$q_i^* = \frac{S(a) - (1-a)c}{n+1}, \quad \pi_i^* = \Big(\frac{S(a) - (1-a)c}{n+1}\Big)^2.$$

and obtain the following comparative statics:

$$\frac{\partial q_i^*}{\partial a} = \frac{S'(a) + c}{n+1}, \quad \frac{\partial \pi_i^*}{\partial a} = 2q_i^* \frac{S'(a) + c}{n+1}.$$

Define the inflection point for the focal occupation as the unique solution $a^* \in (0,1)$ of the equation $S'(a) + c = 0$. We obtain the following proposition.

Proposition 1 (Inflection Point) *q_i^* and π_i^* increase in a when $a < a^*$ but decreases in a when $a > a^*$.*

In other words, for a given occupation, human workers produce more services and earn more as AI improves, so long as the AI performance is below the inflection point of that occupation. However, once AI performance crosses the inflection point, further improvement of AI results in reduced production and lower earnings by human workers. We refer to this model prediction as the *inflection point conjecture*.

Clearly, different occupations have different inflection points. When a new technology leap shocks the CIS, workers in an occupation affected by the surface shock may experience increased transaction volume and greater earnings if AI has not crossed the inflection point.

But if AI has crossed the inflection point before the shock or after the shock, assuming the shock is sufficiently significant, then workers will experience a decreased transaction volume and lower earnings. We also present an alternative competition model (Bertrand model) in Appendix B, which demonstrates the same inflection point conjecture. To test the inflection conjecture, we use the surface shock caused by the launch of ChatGPT.

3.3 Hypotheses

To develop testable hypotheses, we focus on two occupations. The first is the occupation of "Translation". From the perspective of an AI, a translation task involves taking a sequence of words from a source language as input and producing a corresponding sequence of words in a target language as output. The effectiveness of machine translation hinges on its ability to accurately reflect aspects like linguistic precision, semantic consistency, and cultural awareness. Early machine translation tools in the 1950s were primarily rule-based, functioning like a bilingual dictionary with a set of linguistic rules (Goutte 2009, Pestov 2018). They often produced awkward and barely readable translations, far from being practical for real-world use. Therefore, this early form of machine translation was significantly below the minimal intelligence surface required for effective translation jobs. In the 1990s, a paradigm shift occurred with the advent of statistical machine translation tools. These tools departed from rigid rule-setting and instead analyzed text patterns in two languages, learning to translate based on observed statistical patterns, such as correlations between words, phrases, syntax, etc. This resulted in a substantial improvement in translation quality and was incorporated into tools like Google Translate. Despite this progress, their accuracy still fell short of human translators, often producing translations that required further verification and correction for final use. Nevertheless, they were useful to human translators by enhancing their productivity. Since 2015, deep learning architectures such as recurrent neural network (RNN) and transformer (Vaswani et al. 2017) further boosted AI performance on machine translation tasks. In particular, the emergence of the transformer architecture, which powers most LLM models including ChatGPT, ushered in a new era for natural language processing. The underlying attention mechanisms allow the architecture to more accurately and comprehensively identify and exploit the statistical regularities within natural language text, which is particularly important for machine translation. In fact, the transformer architecture was originally proposed to tackle the challenge of machine translation. Of course, a powerful architecture itself is not sufficient for good AI performance in translation, we also need sufficient statistical and computational resources. Fortunately, the Internet has accumulated a large amount of text data, which, paired with high-performance GPUs, finally led to a leap in AI's ability to master natural languages.

The capability of LLMs to manage a wide range of translation tasks has been thoroughly validated in real-world settings (Popel et al. 2020). ChatGPT, for instance, can produce translations of a quality comparable to those done by human translators or commercial translation products (OpenAI 2023). Modern AI's impressive ability to translate natural language is crucial in assisting businesses to overcome language barriers and expand their global reach (George and George 2023), by translating product descriptions and business contracts and addressing customer inquiries from diverse linguistic backgrounds. ChatGPT also demonstrates its utility by offering automated translation of educational content, including textbooks and lectures (Kasneci et al. 2023).

Based on the analyses and evidence above, we believe that for translation, AI has crossed the inflection point and its relation with this occupation has entered the substitution phase for a sufficient number of potential employers, at least after the launch of ChatGPT. By the inflection point conjecture, we propose the following hypothesis for the empirical test.

Hypothesis 1*: *The launch of ChatGPT makes human translators worse off in terms of transaction volume and total earnings.

The second occupation we focus on is "Web Development", for which AI has yet to reach the inflection point. Unlike translation, the occupation of web development involves a variety of tasks, including both front-end and back-end development, and requires skills for both low-level implementation and high-level design. These multifaceted tasks demand a comprehensive skill set, such as programming proficiency, problem-solving skills, debugging, systematic planning, and design expertise. For instance, a front-end developer needs to use different programming languages such as JavaScript, CSS, and HTML to build conceptually distinct components and connect them to create effective and efficient user interfaces. Similarly, a back-end developer needs to interact with the front-end and various databases (e.g., MySQL and MongoDB) as well as other existing systems. The large number of distinct components and the myriad possibilities of implementing and combining them result in very high overall statistical complexity for web development. Moreover, in contrast to the enormous amount of text data publicly available for AI to learn natural languages, due to security and intellectual property concerns, there are few complete website source code publicly available for AI to learn the design and implementation of modern dynamic websites. The combination of high statistical complexity and insufficient statistical resources severely constrains AI performance for web development. In addition, unlike natural languages that are relatively static, techniques for web development evolve much more rapidly, further limiting AI's potential to substitute human developers.

Despite its limitations, AI can quickly build code blocks for human developers to use which can greatly improve their productivity, especially for less experienced developers. In a recent study, Peng et al. (2023) found that by using GitHub Copilot, a tool powered by OpenAI's generative AI model, web developers can implement an HTTP server in Javascript 55.8% faster than developers without access to this AI tool. Essentially, AI can assist developers by automating mundane tasks such as implementing commonly used functionalities, code refactoring, and error warnings. In particular, ChatGPT offers substantial assistance to web developers through the generation, recommendation, explanation, and debugging of code blocks. These enhanced capabilities have led to many online tutorials on using Chat-

GPT for web development [3]. However, like earlier AI tools, ChatGPT cannot fully automate the development of dynamic websites based on user needs. Even the accuracy and robustness of its code outputs require verification and correction by human developers. In terms of design, ChatGPT can be used to brainstorm ideas and come up with an outline with the necessary pages. However human developers still need to step in to ensure the efficiency, scalability, and security of the design. Overall, while ChatGPT offers significant support in web development, surpassing other AI tools in many aspects, especially for less experienced developers, it serves mostly as a helper for speeding up the development rather than a substitute for human developers [4].

Based on the analyses and evidence above, we believe AI has not crossed the inflection point for the occupation of web development since its relation with this occupation is still in the honeymoon phase for most, if not all employers. Based on the inflection point conjecture, we propose the following hypothesis for empirical testing:

***Hypothesis 2*: The launch of ChatGPT makes human web developers better off in terms of transaction volume and total earnings.**

For empirical identification, we need an occupation as the control group which is not affected by the launch of ChatGPT. We choose the occupation of "Construction Design". Jobs within this occupation often require specialized knowledge and confidentiality to some extent. Its statistical regularity between job input and output, if detectable, primarily exists in the related cartographic work or in complex text data. Researchers in the architecture, engineering, and construction (AEC) sector have emphasized that construction design work has been slow to digitize due to its fragmented structure and reliance on specialized skills (Wong et al. 2018). Additionally, the workforce in the construction design sector is relatively

[3] See, for example, https://www.hostinger.com/tutorials/build-website-with-chatgpt/

[4] As the CEO of Meetanshi recently pointed out, "*ChatGPT or any other AI tool will not replace human developers; but can significantly increase their overall productivity.*" For details, see https://meetanshi.com/blog/will-chatgpt-replace-developers/.

small, further complicating the acquisition of relevant data. Limited data availability makes it difficult for AI to learn sufficiently well to be useful. Although there are some efforts to integrate ChatGPT into construction design software like 3D Max and Revit, practitioners report that these initiatives are still in the conceptual phase, with practical implementation for independent projects still a distant prospect [5]. As a result, the occupation of construction design is considered one of the least impacted by ChatGPT (Eloundou et al. 2023), making it a suitable control for our identification purpose.

4 Empirical Evidence

4.1 Data and Background

ChatGPT (i.e., Chat Generative Pre-trained Transformer) is an LLM-based chatbot, which was developed by OpenAI and launched on November 30, 2022. ChatGPT's launch marked a pivotal moment, as its text-generation capabilities and conversation-based design led to widespread use in various applications among both individuals and organizations. In addition, ChatGPT is the first popular generative AI tool that has become mainstream to the general public. Statistics indicate that the introduction of this new GPT version is expected to impact approximately 80% of the U.S. workforce (Eloundou et al. 2023). We hence leverage the launch of ChatGPT as an exogenous shock to test the inflection point conjecture.

Specifically, we use a popular online labor platform as our empirical context. Jobs on this platform cover a large variety, such as translation, writing, data science, construction, and physical science, which allows us to examine how AI influences different job categories, as interpreted in our conceptual framework. The jobs posted on this platform can be classified into two types depending on their price specification, i.e., fixed-price jobs and hourly-rated

[5] https://www.autodesk.com/products/fusion-360/blog/current-and-future-design-automation-tools/

jobs. The fixed-priced job openings provide the total amount of compensation for the job, while the hourly-rated job openings provide a guide for the hourly price of the job and the estimated duration of the job. After a job is posted, workers who are interested can submit their proposals to the employer. Subsequently, the employer will review these applications and work proposals to select appropriate workers for the job vacancies. Upon completion of the work, the employer releases the payment due, and they can provide ratings and reviews for the worker based on the quality of the work. Under the worker profile, we are able to retrieve the history of all his/her jobs for analysis.

This platform aligns well with our research objectives for two primary reasons. First, the platform has a well-structured job classification system that spans from a broad "category" to a narrower "sub-category" and more granular "specialties". In alignment with this, workers are categorized into particular focal online labor markets (OLMs) by the platform, based on the majority of jobs they have taken. This detailed system offers a clear portrayal of jobs necessitating specialized skills and corresponding human labor in OLMs. This allows us to obtain worker-level transaction histories related to distinct job categories. Secondly, the platform offers a wealth of information about workers, including their names, skills, location, default hourly rate, language proficiency, educational background, and so on. Additionally, the platform grants full access to the entire work history of its workers, including specifics such as job titles, received ratings, job start and end dates, job prices, and comments from employers. This enables us to accurately measure the acceptance time, completion time, and payment for jobs undertaken by workers since their registration.

We identified workers engaging in each specific OLM by the "specialties" classification provided on the platform. In compliance with platform policy, we retrieved up to 5,000 workers for each specialty. Table A-1 in Appendix A provides a summary of the OLM classification for different values of "specialties". In total, we obtained profiles and work histories of 6,743 workers belonging to the Construction Design OLM, 7,582 workers belonging to the Translation OLM, and 15,000 workers belonging to the Web Development OLM. We then removed the inactive workers who had not accepted any job prior to November 1, 2022, and aggregated the data at the worker level on a monthly basis. A worker within a specific market may possess multiple skills enabling them to engage in jobs beyond their primary OLM. In this paper, we define jobs aligned with workers' primary labor market as "focal jobs", while others as "non-focal jobs". The proportions of focal jobs to total jobs accepted by all workers are 64.1% for the Translation OLM, 81.3% for the Web Development OLM, and 67.2% for the Construction Design OLM.

4.2 Research Design & Measurements

The goal of our empirical study is to analyze the impact of AI on different occupations, we hence focus on each worker's focal jobs within each OLM in the analysis. To further ensure the observed effects occurred after ChatGPT's launch, all measurements were constructed based on the focal jobs accepted within a given month, rather than those completed. We excluded data from November and December 2022 to account for potential pre-launch impacts of ChatGPT and holiday effects. Hence, the study's time frame spans six months before and after the shock, from May 1, 2022, to June 30, 2023. Table 1 provides details on the key variables and their definitions. Table 2 reports the descriptive statistics of workers from the three OLMs for the main analyses.

We test the hypotheses using the following two-way fixed-effect difference-in-differences

Table 1: Definitions of Key Variables

Variables	**Definitions**
$Fjobearn_{it}$	The total earnings of focal jobs accepted in month t by worker i
$Fjobnum_{it}$	The number of focal jobs accepted in month t by worker i
$Fjobratio_{it}$	The ratio of accepted focal jobs to all accepted jobs in month t by worker i
$Fjobprice_{it}$	The average price per focal job accepted in month t by worker i
$Fjobrating_{it}$	The average rating of focal jobs accepted in month t by worker i
$Fhourprice_{it}$	The average hourly rate of focal jobs accepted in month t by worker i
$Tenure_{it}$	The number of months since worker i's registration up to month t

Table 2: Descriptive Statistics of Key Variables

Measure	Count	Mean	Min	Max	Std. dev.
$Fjobearn_{it}$	273672	468.255	0.000	590183.690	4684.211
$Fjobnum_{it}$	273672	0.386	0.000	45.000	1.128
$Fjobratio_{it}$	273672	0.195	0.000	1.000	0.384
$Fjobprice_{it}$	53594	1747.721	1.000	590183.690	9273.679
$Fjobrating_{it}$	35296	4.872	1.000	5.000	0.450
$Fhourprice_{it}$	23300	27.377	3.000	357.500	20.594
$Tenure_{it}$	273672	37.758	0.000	226.000	34.733

Note: If worker i does not accept any focal jobs in month t, $Fjobprice_{it}$ would be recorded as a null value, and $Fjobratio_{it}$ would be recorded as zero.

(DID) model for identification where the unit of analysis is at the worker-month level.

$$Y_{it} = \beta_0 + \beta_1 \times ChatGPT_{it} + \beta_2 \times Tenure_{it} + \eta_i + \tau_t + \epsilon_{it} \tag{1}$$

In Equation (1), i and t index worker and month, respectively. The dependent variable Y_{it} measures worker i's transaction volume or total earnings in the focal OLM during month t. For the transaction volume, we use $\log(Fjobnum_{it})$ to measure the log-transformed number of focal jobs worker i accepts in month t. We also use a relative measure, $Fjobratio_{it}$, which is the ratio of accepted focal jobs to the total number of accepted jobs by worker i in month t. For earnings, we use $\log(Fjobearn_{it})$ to measure worker i's total earnings from focal jobs in month t. The explanatory variable of interest is the binary variable $ChatGPT_{it}$ (i.e., the DID estimator $Treat_i \times After_t$) which equals 1 if worker i's main job category is the treated job category and the transaction activities under investigation occurred after the launch of ChatGPT. Otherwise, the binary variable $ChatGPT_{it}$ equals 0. η_i captures the worker fixed effect, while τ_t captures the time fixed effect. We also include a control for worker tenure, i.e., $Tenure_{it}$, measured by the number of months up to month t since worker i's registration.

To ensure workers in the treated and control groups are comparable, we used propensity score matching (PSM) to improve the sample balance by accounting for workers' tenure and experience (i.e., the total number of accepted focal jobs), wages (i.e., average price and average hourly rate of focal jobs) and quality of work (i.e., average rating of focal jobs).

All variables for matching were calculated from the work history data before the launch of ChatGPT. We adopted a 1:1 nearest-neighbor matching strategy at the worker level and excluded observations falling outside of the common support region (Caliendo and Kopeinig 2008).

4.3 Main Results

Our first analysis tests Hypothesis 1, by examining the effect of ChatGPT on translation workers, using the workers in the Construction Design market as the control group. After matching with a caliper value of 2.05×10^{-4}, we obtain 2,295 workers. Table A-2 in Appendix A reports the balance test results before and after the matching.

Table 3 reports the DID estimation results. Overall, we find strong support for Hypothesis 1. More specifically, in column (1) which corresponds to the dependent variable of $\log(Fjobnum_{it})$, we find the coefficient of $ChatGPT_{it}$ negative and statistically significant, suggesting a decrease in the absolute number of focal jobs accepted by workers after ChatGPT's launch. In terms of magnitude, the transaction volume dropped by 7.4% ($= 1 - e^{-0.077}$) on average. In column (2) which corresponds to the dependent variable $Fjobratio_{it}$, the coefficient of $ChatGPT_{it}$ is also negative (-0.054) and statistically significant, indicating that workers accept fewer focal jobs in the relative term as well. In column (3) which corresponds to the dependent variable $\log(Fjobearn_{it})$, the coefficient of $ChatGPT_{it}$ is also negative and statistically significant, suggesting a decrease in worker's earnings from focal jobs after ChatGPT's launch, by 30.2% ($= 1 - e^{-0.360}$) on average. This result also indicates a significant decline in the platform's revenue from the Translation sector.

In summary, these estimation results support Hypothesis 1 and suggest that AI has crossed the inflection point for translation jobs. It seems that the launch of ChatGPT has shocked the CIS significantly enough for translation jobs.

Our second analysis tests Hypothesis 2, by examining the effect of ChatGPT on web developers, again using the workers in the Construction Design market as the control group.

Table 3: Effect of ChatGPT on Translation Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
ChatGPT	-0.077***	-0.054***	-0.360***
	(0.015)	(0.011)	(0.074)
Observations	27540	27540	27540
N	2295	2295	2295
Adjusted R^2	0.482	0.270	0.336

Note: (1) *p<0.1, **p<0.05, ***p<0.01; (2) Clustered standard errors are in the parentheses; (3) We control for time fixed effect, worker fixed effect and worker tenure. Unless otherwise noted, the same specifications are applied in the subsequent tables.

After matching with a caliper value of 3.1×10^{-5}, we obtained data for 2,361 workers. Table A-3 in Appendix A reports the balance test results.

Table 4 reports the DID estimation results. In contrast to the results for translation workers, we find the opposite effects. Specifically, we find a 7.3% ($= e^{0.070} - 1$) increase on average in transaction volume for web developers after ChatGPT became available. This is also true in terms of relative transaction volume, as is suggested by the estimated coefficient of $ChatGPT_{it}$ in column (2), which corresponds to the dependent variable $Fjobratio_{it}$. Furthermore, in column (3), the estimated coefficient of $ChatGPT_{it}$ corresponding to the dependent variable $\log(Fjobearn_{it})$ is also positive and statistically significant, with a magnitude of nearly 60.0%($= e^{0.470} - 1$). Therefore, we find strong evidence supporting Hypothesis 2. As we discussed previously, the availability of ChatGPT is unlikely to automate the process of web development but can improve a developer's productivity. Hence, AI and web development jobs are likely to be in the honeymoon phase for most projects.

4.4 Parallel Trend Assumption

The parallel trend assumption and the no-anticipation assumption are key to the validity of DID analysis. To provide empirical support, we conduct a lead-and-lag test, by estimating

Table 4: Effect of ChatGPT on Web Development Jobs

	Variables		
	(1) log(Fjobnum)	(2) Fjobratio	(3) log(Fjobearn)
ChatGPT	0.070***	0.057***	0.470***
	(0.013)	(0.011)	(0.077)
Observations	28332	28332	28332
N	2361	2361	2361
Adjusted R^2	0.386	0.237	0.278

Note: (1) *p<0.1, **p<0.05, ***p<0.01

the following relative-time model:

$$Y_{it} = \beta_0 + \sum\nolimits_{\sigma=-6}^{\sigma=5} \beta_\sigma \times Rel\ Time_\sigma \times Treat_i + \beta_2 \times Tenure_{it} + \eta_i + \tau_t + \epsilon_{it} \qquad (2)$$

In Equation (2), $Rel\ Time_\sigma$ is a binary variable, which represents the relative month σ to the launch month of ChatGPT. $Treat_i$ is 1 if worker i is in the treated occupation, and is 0 otherwise. We omit the first month prior to ChatGPT's launch which serves as the baseline period. The set of coefficients β_σ indicates whether differences between workers in treated and controlled OLMs exist before ChatGPT's launch ($\sigma < 0$) and how the estimated effects change over time afterward ($\sigma \geq 0$).

We report the results in Table 5 and Table 6, for translation jobs and web development jobs, respectively. For all dependent variables and both job categories, we find that the estimated coefficients β_σ are insignificant before ChatGPT's launch, which are consistent with our identification assumptions. The effects on $\log(Fjobnum_{it})$ and $\log(Fjobearn_{it})$ become significantly negative or positive after ChatGPT's launch in each analysis. Interestingly, we find that the negative effect of ChatGPT on the transaction volume of translation jobs seems to strengthen over time, especially after March 2023, possibly due to the release of GPT-4 on March 14, 2023. In contrast, the positive effect of ChatGPT for web development is more stable. One explanation is that while web developers can immediately take advantage of ChatGPT, employers may need some time to assess the feasibility of substituting ChatGPT for translators.

Table 5: Relative-time Model: Effect of ChatGPT on Translation Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
Rel Time (t-6)	-0.017	-0.006	-0.058
	(0.028)	(0.024)	(0.152)
Rel Time (t-5)	-0.009	0.004	0.014
	(0.026)	(0.024)	(0.150)
Rel Time (t-4)	0.007	0.012	0.041
	(0.026)	(0.024)	(0.152)
Rel Time (t-3)	-0.002	0.015	-0.004
	(0.026)	(0.025)	(0.152)
Rel Time (t-2)	0.017	0.034	0.175
	(0.023)	(0.023)	(0.143)
Rel Time (t)	-0.064**	-0.030	-0.246*
	(0.026)	(0.023)	(0.143)
Rel Time (t+1)	-0.057**	-0.038	-0.300**
	(0.025)	(0.024)	(0.140)
Rel Time (t+2)	-0.066***	-0.045**	-0.342**
	(0.026)	(0.022)	(0.135)
Rel Time (t+3)	-0.085***	-0.048**	-0.386***
	(0.025)	(0.022)	(0.137)
Rel Time (t+4)	-0.099***	-0.044*	-0.321**
	(0.026)	(0.023)	(0.138)
Rel Time (t+5)	-0.093***	-0.062***	-0.395***
	(0.027)	(0.023)	(0.142)
Observations	27540	27540	27540
N	2295	2295	2295
Adjusted R^2	0.482	0.270	0.335

Note: (1) *p<0.1, **p<0.05, ***p<0.01; (2) Clustered standard errors are in the parentheses; (3) We control for time fixed effect, worker fixed effect and worker tenure.

5 Additional Job Categories

To further test the inflection conjecture, we consider three more job categories: writing, machine learning, and physical sciences. We did not use these job categories for the main analyses because of the following considerations. Although we have reasons to believe that AI has crossed the inflection point for writing, especially after the launch of ChatGPT, it

Table 6: Relative-time Model: Effect of ChatGPT on Web Development Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
Rel Time (t-6)	0.016	0.014	0.143
	(0.026)	(0.025)	(0.167)
Rel Time (t-5)	-0.017	-0.027	-0.155
	(0.025)	(0.024)	(0.165)
Rel Time (t-4)	0.001	-0.010	-0.017
	(0.025)	(0.025)	(0.164)
Rel Time (t-3)	0.004	0.010	0.039
	(0.026)	(0.026)	(0.171)
Rel Time (t-2)	0.005	0.003	0.063
	(0.023)	(0.024)	(0.164)
Rel Time (t)	0.068***	0.043*	0.377**
	(0.025)	(0.024)	(0.159)
Rel Time (t+1)	0.083***	0.076***	0.558***
	(0.025)	(0.024)	(0.162)
Rel Time (t+2)	0.079***	0.061**	0.517***
	(0.025)	(0.024)	(0.165)
Rel Time (t+3)	0.042*	0.032	0.321**
	(0.024)	(0.024)	(0.162)
Rel Time (t+4)	0.077***	0.070***	0.555***
	(0.025)	(0.024)	(0.162)
Rel Time (t+5)	0.079***	0.050**	0.565***
	(0.024)	(0.023)	(0.154)
Observations	28332	28332	28332
N	2361	2361	2361
Adjusted R^2	0.386	0.237	0.278

Note: (1) *p<0.1, **p<0.05, ***p<0.01; (2) Clustered standard errors are in the parentheses; (3) We control for time fixed effect, worker fixed effect and worker tenure.

is not as clear-cut as translation. For the job category of machine learning, we believe AI has not yet crossed the inflection point, much like web development. However, because machine learning is a branch of AI, the launch of ChatGPT may have a spillover effect on the demand for machine learning jobs. Finally, the job category of physical sciences often requires sophisticated reasoning which statistical AI is not particularly good at. So, we believe AI has not crossed the inflection point for this category either. However, compared to other markets, this is a thin market without many workers and jobs on the platform.

Despite the limitations of these job categories, we report our empirical analyses here as additional supporting evidence.

5.1 Job Category: Writing

Large language models, especially ChatGPT, excel at writing. Although not as obvious as translation, existing evidence suggests that AI performance likely has crossed the inflection point for this job category as well (OpenAI 2023). To test the inflection point conjecture for this job category, we collected profiles and work histories of 22,184 workers from the "Content, Professional, and Business Writing" classifications on the platform. Again, we use construction design as our control group and use PSM (see Table A-4 in Appendix A for balance check results) to construct a matched sample. Table 7 reports the estimation results from the two-way fixed-effect DID model. Similar to the findings from the job category of translation, we find a negative and statistically significant effect of ChatGPT's launch on the transaction volume and earnings from writing jobs, thereby supporting the inflection point conjecture.

In terms of magnitude, we do find that workers in the Writing OLM experience less of a decline, with transaction volume down 4.4%, and earnings down 19.2%, compared to workers in the Translation OLM. This makes sense because compared to translation, writing jobs involve more tasks that require creativity and critical thinking. As a result, the region of CIS for writing jobs is less shocked by ChatGPT than the region of CIS for translation jobs.

Table 7: Effect of ChatGPT on Writing Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
ChatGPT	-0.045***	-0.035***	-0.213***
	(0.010)	(0.009)	(0.055)
Observations	55308	55308	55308
N	4609	4609	4609
Adjusted R^2	0.370	0.235	0.287

Note: (1) *p<0.1, **p<0.05, ***p<0.01

5.2 Job Category: Machine Learning

Like web development, the job category of machine learning requires problem-solving skills and proficiency in programming, albeit with different languages (e.g., Python and R). Because AI's ability to write complete code is still limited, even after the availability of ChatGPT, we expect AI to be more of an assistant to machine learning developers, at least for most non-trivial jobs. Hence, we posit that AI performance has not crossed the inflection point for machine learning jobs. On the other hand, the availability of ChatGPT should still greatly benefit machine learning developers by providing code snippets and helping with debugging.

To test the inflection point conjecture, we collected profiles and job histories of 4,069 workers from the "Machine Learning" classifications on the platform. Utilizing the same empirical strategy, we found statistically significant and positive effects of ChatGPT's launch on transaction volume and total earnings, as reported in Table 8.

Interestingly, we find that the positive effect is less strong in terms of magnitude than that of web development, especially for the transaction volume. Given that our estimate for this job category is likely a slight overestimation due to the potential spillover effect of ChatGPT on machine learning, the real difference in the effect of ChatGPT should be greater between these two job categories. To understand this discrepancy, consider the differences between these two job categories in terms of statistical complexity and statistical resources. As we discussed earlier, a typical web development job requires the interoperation of several distinct components implemented in different programming languages, involving multiple systems and asynchronous event handling. A typical machine learning project, on the other hand, can be completed using just one programming language (e.g., Python or R) in a linear structure. So, the statistical complexity is likely much higher for tasks associated with web development than tasks associated with machine learning. Moreover, there is a greater volume of machine learning code publicly available than website source code, due to security and intellectual property concerns of the latter. Therefore, compared with web development, the CIS region for machine learning is more likely to be above the minimal intelligence surface for many potential employers. As a result, the availability of ChatGPT will likely have seen a greater substitution for machine learning than for web development which led to a less pronounced positive effect for machine learning.

Table 8: Effect of ChatGPT on Machine Learning Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
ChatGPT	0.059***	0.052***	0.382***
	(0.014)	(0.012)	(0.081)
Observations	24180	24180	24180
N	2015	2015	2015
Adjusted R^2	0.306	0.205	0.246

Note: (1) *p<0.1, **p<0.05, ***p<0.01

5.3 Job Category: Physical Science

Beyond its proficiency in text generation tasks like writing and programming, ChatGPT's capabilities in mathematical and scientific fields have attracted considerable academic interest. Empirical evidence shows a substantial enhancement in ChatGPT's knowledge in various subjects such as mathematics and chemistry (OpenAI 2023, Eloundou et al. 2023). However, the notable achievements of ChatGPT in mathematical and scientific disciplines primarily stem from its ability to learn statistical regularities from extensive examples, rather than a true mastery of mathematical and logical reasoning, as is evidenced by the phenomenon of hallucination. So, it's reasonable to believe that AI performance has yet to cross the inflection point in the realm of physical sciences if it can at all. Nevertheless, the availability of ChatGPT does make human workers more productive by serving as an aid that can quickly generate a possible solution.

We collected profiles and job histories of 1,632 workers from the "Physical Sciences" classifications on the platform. Utilizing the same empirical strategy, we estimated and found positive and statistically significant effects of ChatGPT's launch on each dependent variable, as reported in Table 9. These findings further support the inflection point conjecture.

In terms of magnitude, the effects are smaller than those for the job of web development and those for the job of machine learning. We believe the reasons are twofold. First, unlike large projects in machine learning and web development, these jobs on the platform are usually smaller, making it more likely for many potential employers to substitute AI for human workers. Second, AI limitations on these jobs constrain the level of assistance it can provide to human workers. Both mechanisms lead to a small effect from ChatGPT on this job category, although the small sample size also has led to less precise estimations of the effect size.

Table 9: Effect of ChatGPT on Physical Sciences Jobs

	Variables		
	(1)	(2)	(3)
	log(Fjobnum)	Fjobratio	log(Fjobearn)
ChatGPT	0.044***	0.036**	0.194**
	(0.016)	(0.016)	(0.096)
Observations	12276	12276	12276
N	1023	1023	1023
Adjusted R^2	0.364	0.226	0.285

Note: (1) *p<0.1, **p<0.05, ***p<0.01

6 Conclusions

This paper contributes both theoretically and empirically to our understanding of AI's implications on jobs. On the theoretical front, we analyzed a three-phase relation between AI and jobs and proposed the inflection point conjecture that contrasts the effect of AI progress on workers in two distinct stages. Before AI performance crosses the inflection point of an occupation, human workers always benefit from improvements in AI, but after AI performance crosses the inflection point, human workers become worse off whenever AI performance improves. Our analysis is rooted in three elements of thoughts. First, we proposed a visual framework to think about AI performance through the lens of four factors:

task learnability — an inherent property of each task consisting of both statistical and computational aspects, statistical and computational resources, and learning techniques which glue together the other three factors. Second, by considering an occupation as a task set, we compared the current intelligence surface with the minimal intelligence surface over the task set, which naturally leads to three phases of the AI-jobs relation: decoupled, honeymoon, and substitution. Third, the evolving AI-jobs relation is convoluted by the heterogeneity among potential employers in terms of their AI literacy and the acceleration of technology adoption as it matures. Despite its qualitative nature, we believe this theoretical framework is a useful instrument to systematically examine the underlying forces driving the AI revolution and to understand its far-reaching implications in the future.

On the empirical side, we tested the inflection point conjecture using data from a large online labor platform. Empirical analyses based on two of its markets suggest that AI's relation with the occupation of translation has passed the inflection point, but AI's relation with the occupation of web development has not. These findings have important implications for workers of these two occupations. As AI performance keeps improving, we expect workers in the translation industry to continue to be hurt by these improvements while web developers continue to benefit from them. Such an occupation-wise inflection point analysis can help us evaluate the concurrent relation between AI and various occupations, and offer insights on the prospect of these occupations as AI march steadily forward. We believe this is particularly valuable as the society prepares the next generation of workers in the age of AI revolution. As such, the main limitation of the current paper is the small number of occupations empirically examined. Another limitation is the fact that our data is from one platform. Given the potentially large disruption of AI on employment, more studies on more occupations using data from different platforms are urgently needed.

References

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *J. Econom. Perspect.*, 33(2):3–30.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *J. Econom. Perspect.*, 33(2):31–50.

Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing inequality: Have computers changed the labor market? *Quart. J. Econom.*, 113(4):1169–1213.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quart. J. Econom.*, 118(4):1279–1333.

Barach, M. A., Golden, J. M., and Horton, J. J. (2020). Steering in online markets: The role of platform incentives and credibility. *Management Sci.*, 66(9):4047–4070.

Benson, A., Sojourner, A., and Umyarov, A. (2020). Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Sci.*, 66(5):1802–1825.

Bresnahan, T. F., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quart. J. Econom.*, 117(1):339–376.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *J. Econom. Surveys*, 22(1):31–72.

Chen, Z. and Chan, J. (2023). Large language model in creative work: The role of collaboration modality and user expertise. Working Paper, University of Minnesota.

Chwelos, P., Ramirez, R., Kraemer, K. L., and Melville, N. P. (2010). Research note—Does technological progress alter the nature of information technology as a production input? New evidence and new results. *Inform. Systems Res.*, 21(2):392–408.

Dewan, S. and Min, C.-k. (1997). The substitution of information technology for other factors of production: A firm level analysis. *Management Sci.*, 43(12):1660–1675.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *Preprint, submitted March 17, https://doi.org/10.48550/arXiv.2303.10130.*

George, A. S. and George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1):9–23.

Goutte, C. (2009). *Learning machine translation.* MIT Press, London.

Hong, Y., Wang, C., and Pavlou, P. A. (2016). Comparing open and sealed bid auctions: Evidence from online labor markets. *Inform. Systems Res.*, 27(1):49–69.

Horton, J. J. (2010). Online labor markets. In Saberi, A., editor, *Internet and Network Economics*, pages 515–522. Springer, Berlin Heidelberg.

Horton, J. J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *J. Labor Econom.*, 35(2):345–385.

Huang, N., Burtch, G., Hong, Y., and Pavlou, P. A. (2020). Unemployment and worker participation in the gig economy: Evidence from an online labor market. *Inform. Systems Res.*, 31(2):431–448.

Kahneman, D. (2013). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kanat, I., Hong, Y., and Raghu, T. (2018). Surviving in global online labor markets for IT services: A geo-economic analysis. *Inform. Systems Res.*, 29(4):893–909.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., and Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274.

Kokkodis, M. (2023). Adjusting skillset cohesion in online labor markets: Reputation gains and opportunity losses. *Inform. Systems Res.*, 34(3):1245–1258.

Kokkodis, M. and Ipeirotis, P. G. (2023). The good, the bad, and the unhirable: Recommending job applicants in online labor markets. *Management Sci.*, 69(11):6969–6987.

Leung, M. D. (2018). Learning to hire? Hiring as a dynamic experiential learning process in an online market for contract labor. *Management Sci.*, 64(12):5651–5668.

Lin, M., Liu, Y., and Viswanathan, S. (2018). Effectiveness of reputation in contracting for customized production: Evidence from online labor markets. *Management Sci.*, 64(1):345–359.

Liu, J., Xu, X., Li, Y., and Tan, Y. (2023). "Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets. *Preprint, submitted August 9, https://doi.org/10.48550/arXiv.2308.05201*.

Lu, S. F., Rui, H., and Seidmann, A. (2018). Does technology substitute for nurses? Staffing decisions in nursing homes. *Management Sci.*, 64(4):1842–1859.

Lysyakov, M. and Viswanathan, S. (2022). Threatened by AI: Analyzing users' responses to the introduction of AI in a crowd-sourcing platform. *Inform. Systems Res.*, 34(3):1191–1210.

OpenAI (2023). Gpt-4 technical report. *Preprint, submitted March 15, https://doi.org/10.48550/arXiv.2303.08774*.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from Github Copilot. *Preprint, submitted February 13, https://doi.org/10.48550/arXiv.2302.06590*.

Pestov, I. (2018). A history of machine translation from the Cold War to deep learning. Retrieved November 24, 2023, from https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5/.

Popel, M., Tomkova, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., and Žabokrtskỳ, Z. (2020). Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11(1):4381.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. (2017). Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017).*

Wies, N., Levine, Y., and Shashua, A. (2023). The learnability of in-context learning. *Preprint, submitted March 14, https://doi.org/10.48550/arXiv.2303.07895.*

Wong, J. K. W., Ge, J., and He, S. X. (2018). Digitisation in facilities management: A literature review and future research directions. *Autom. Constr.*, 92:312–326.

Xue, M., Cao, X., Feng, X., Gu, B., and Zhang, Y. (2022). Is college education less necessary with AI? Evidence from firm-Level labor structure changes. *Journal of Management Information Systems*, 39(3):865–905.

Yu, Y., Yang, Y., Huang, J., and Tan, Y. (2023). Unifying algorithmic and theoretical perspectives: Emotions in online reviews and sales. *MIS Quart.*, 47(1):127–160.

Zhang, D., Cheng, Z., Mohammad, H. A. Q. H., and Nault, B. R. (2015). Research commentary—Information technology substitution revisited. *Inform. Systems Res.*, 26(3):480–495.

Zhang, D. D., Peng, G., Yao, Y., and Browning, T. R. (2023). Is a college education still enough? The IT-labor relationship with education level, task routineness, and artificial intelligence. *Inform. Systems Res. Forthcoming.*