Empowering Inclusive Work

Yanyou Chen
University of Toronto

Mitchell Hoffman
UC Santa Barbara and NBER

Huilan Xu
Zhejiang University

Zhe Yuan
Zhejiang University

November 2025

Abstract

Can AI improve workplace outcomes for workers with disabilities? We examine the relative performance of hearing-impaired workers on one of China's largest food-delivery platforms. Pre-AI, disabled workers are slightly slower than non-disabled workers and have worse customer ratings, although they supply more hours to the platform and are much less likely to quit. Midway through our data, the platform suddenly introduces an AI-based intelligent outbound call system designed to improve communication for hearing-impaired workers. AI significantly increases the productivity of disabled workers, especially on tasks involving interaction with customers; reduces late deliveries and negative customer reviews; and increases labor supply and retention. AI eliminates over half of the disability hourly pay gap and substantially increases the profitability of disabled workers for the platform.

Keywords: Artificial intelligence, Disability, Productivity, Gig workers
JEL Codes: M50, J14
Preliminary — work in progress. Please do not cite without permission.

1 Introduction

Around the world, workers with a disability face challenges in the labor market (Ameri et al., 2018; Baldwin and Johnson, 2006; L'Horty et al., 2022). They have lower rates of employment and face substantial discrimination (Bellemare et al., 2023), and there are large fiscal costs from high use of disability insurance (Autor and Duggan, 2003; Milligan, 2012). Challenges faced by disabled workers are especially acute in middle-income countries like China where government social insurance is often less developed (Banerjee et al., 2024).

Unfortunately, relatively little is known regarding "what works" in improving outcomes for disabled workers. A central policy to increase opportunity for disabled workers is civil rights laws, a key example being the Americans with Disabilities Act (ADA). The ADA prohibits discrimination against disabled workers and mandates employers to provide reasonable accommodations, and similar laws exist in many countries. However, research indicates that the ADA *decreased* the employment of disabled workers, potentially due to increased costs or legal risks for employers (Acemoglu and Angrist, 2001; DeLeire, 2000). Another common policy in many countries (e.g., Austria, Brazil, China) is to impose quotas, requiring larger firms to employ a minimum share of workers with disabilities. However, firms may comply with quotas by hiring disabled workers into lower-paid jobs (Szerman, 2022; Lalive et al., 2013). Rather than imposing compliance costs through civil-rights enforcement or quotas, an alternative is to rely on technologies like artificial intelligence (AI) that directly raise productivity—potentially increasing inclusion while aligning with firms' incentives.

Our study provides the first evidence from a large-scale setting on disabled workers' productivity and the effects of a targeted AI technology. We address two questions:

1. In the same job, how do disabled workers compare to non-disabled workers in terms of productivity, output, profit for firms, and compensation?

2. Can AI tools affect the productivity, compensation, and profitability of disabled workers, and how?

To address these questions, we leverage a unique collaboration with a major food delivery platform in China (hereafter, the "Platform"). The Platform offers an ideal setting for three reasons. First, the Platform operates on an immense scale: it is one of the two dominant food-delivery apps in China, directly engaging about three million active workers and completing hundreds of millions of deliveries per month. Second, and importantly, the Platform follows an open hiring policy: all applicants who meet basic criteria are accepted, with no higher bar or differential screening for disabled workers. Thus, disabled workers are employed under the same conditions as their non-disabled counterparts, allowing for a cleaner comparison

of on-the-job performance (Pallais and Sands, 2016). Third, work on such platforms has become an essential source of employment for disabled individuals in China.

The majority of disabled workers on the Platform are people with hearing impairments. We focus on this group of hearing-impaired workers in our analysis.[1] Focusing on a single, relatively common disability allows us to clearly define the challenges involved and the specific technological solution introduced, while holding constant other factors–an approach that provides concrete insights into how to assist this group. Moreover, hearing impairment is a disability that gig economy platforms around the world are increasingly able to accommodate (e.g., through text-based communications), making our findings broadly relevant.

We use a rich administrative dataset from the Platform, which includes all newly registered disabled workers and a random sample of newly registered non-disabled workers over a three-year period (2021–2023). This dataset allows us to directly compare the job performance of hearing-impaired workers to a large sample of non-disabled workers on multiple dimensions. We examine key performance indicators such as delivery speed, customer satisfaction, number of orders completed, and labor supply (hours worked and attrition/retention). Crucially, in the middle of our observation period, the Platform introduced a novel AI-based "Intelligent Outbound Calling" feature designed specifically to help hearing-impaired workers communicate with customers. This feature—essentially a text-to-speech phone calling system with preset audio prompts—was rolled out suddenly in April 2022 to all hearing-impaired workers. We exploit this event as a natural experiment to evaluate how a targeted AI intervention can impact disabled workers' performance and the Platform's outcomes.

Prior to the new AI tool, disabled (hearing-impaired) workers worked slightly slower than non-disabled workers. This was especially so on the final segment of trips, from restaurant pickup to customer delivery, where communication is most important. Disabled workers are also more likely to receive bad customer ratings than non-disabled workers. Despite this, disabled workers contributed greater labor input than non-disabled workers, working more hours and being less likely to quit. The higher labor supply of disabled workers more than compensated for their small efficiency gap in terms of total output: disabled workers delivered more orders on average per week than non-disabled workers, making them valuable to the platform in the short run despite their limitations. However, because disabled workers work slower, they earn 8% less per hour than non-disabled workers.

After the introduction of the Intelligent Outbound Calling system, the performance of disabled workers improves substantially. Disabled workers' speed of work increases, especially

[1]Hearing impairment affects hundreds of millions of people and is among the most prevalent disabilities, affecting 8% of the Chinese population (Wang et al., 2025) and 13% of US adults (Madans et al., 2021). The share with severe hearing impairment is much lower, but still very substantial.

on the trip segments involving communication where they lagged the most previously. They become less likely to have deliveries and to experience bad ratings. The total number of orders increases, especially among disabled workers who work full independently without the support of other workers. Labor supply and retention further increase. These improvements translate into a higher economic value of disabled workers for the firm, with preliminary evidence suggesting that the gains from the technology likely exceed its costs from the platform's perspective. AI causes the disabled worker pay gap to fall from 8% to 3%, a decrease of over half. The results support that a key mechanism for these effects is an improvement in ability to communicate with customers. Overall, the results suggest that AI substantially increases the return to being a food delivery worker, as well as increases the Platform's returns from employing disabled workers. A limitation of these results is that we currently don't observe behavior off the Platform, but we hope to do a Platform survey to address this issue.

This paper contributes to several literatures, the first one being that on the economics of disability. A large body of work studies the determinants and consequences of disability insurance and other social safety programs for disabled workers (Bound and Burkhauser, 1999; Deshpande, 2016; Autor et al., 2019; Deshpande and Lockwood, 2022). There is also work on disabled worker pay gaps (Forth and Jones, 2025), as well as estimating the impact of civil rights laws and quotas for disabled workers, but relatively little is known about what works in improving labor outcomes for disabled workers.[2] An important exception is Bloom et al. (2025) who argue that the large rise in work from home has helped increase the employment of disabled workers. Much less emphasized in this literature has been the importance of understanding what firms gain from hiring disabled workers. We are not aware of prior work that has been able to compare disabled and non-disabled workers in terms of individual productivity data. Furthermore, we are not aware of prior work estimating the impact of new technologies on disabled workers.

Pre-AI, while disabled workers are less efficient than non-disabled workers, they still appear to be more profitable overall. This highlights the importance of comparing a broad range of outcome variables in evaluating the profitability of disabled workers for firms, raising the intriguing possibility that some firms could overlook the value of disabled workers if they are focused exclusively on some outcome variables. Further, our results indicate that a simple AI technology can be highly valuable in improving the efficiency and overall profitability of disabled workers. While wage subsidies need to be paid each time a disabled worker is hired, new technologies may be more scalable.

[2] Lise et al. (2023) find that the ADA reduced employment for workers with physical work limitations but increased it for those with non-work limitations such as mental illness.

Second, it contributes to work on the economics of AI. Although AI adoption is accelerating, causal evidence on its impact for disadvantaged groups remains scarce. Existing studies examine how AI influences worker performance in general populations, such as call center agents (Brynjolfsson et al., 2025). professionals in laboratory settings (Noy and Zhang, 2023), and software engineers (Cui et al., 2025; Hoffmann et al., 2024). In contrast, our study is the first to evaluate how a targeted AI tool—AI-powered outbound calling—affects productivity, retention, and profitability for disabled workers. We leverage granular administrative data and a quasi-experiment from one of China's largest delivery platforms, providing unique insight into the real-world impact of AI on an underserved population. Work on earlier technologies like computers suggests some role of technology in improving outcomes for disabled workers (Kruse et al., 1996; Krueger and Kruse, 1995), but is not based on administrative data and generally lacks quasi-experimental variation in technology. It also connects to the broader economics literature on IT and productivity in firms (Bresnahan et al., 2002; Bartel et al., 2007).

Third, it contributes to work in personnel and organizational economics on demand for disadvantaged workers. There is growing interest in these fields in worker selection (Oyer and Schaefer, 2011). However, as pointed out in the recent handbook chapter by Hoffman and Stanton (2025), there has been relatively little work on firm returns from disadvantaged workers. One recent exception is Benson et al. (2024), who analyze productivity differences between white workers and workers from under-represented groups in a large US retail firm. There is also work on policies and firm returns in regard to hiring workers with a criminal record (Minor et al., 2018; Cullen et al., 2023). To our knowledge, ours is the first paper in personnel economics to measure what firms gain from hiring disabled workers.

The paper proceeds as follows. Section 2 proposes a theoretical model of worker behavior and firm profits. Section 3 describes the study setting, data, and outcome metrics used to measure productivity differences and the impact of the technology. Section 4 examines the productivity, labor supply, compensation, and profitability of disabled and non-disabled workers prior to the AI intervention and identifies bottlenecks faced by disabled workers. Section 5 analyzes the effects of the AI tool, both overall and heterogeneous treatment effects (still to do). Section 6 discusses a large-scale survey that we've applied for funding to run. The goal is to better understand the work lives of disabled workers on the Platform to elucidate the mechanisms of how they've been affected by AI. Section 7 concludes.

2 Modeling Labor Supply on the Platform

We present a model of labor supply on the Platform to ground our empirical analysis and to make predictions about differences between disabled and non-disabled workers, as well as how disabled worker productivity is affected by AI tools. The model builds on rich ethnographic observations of disabled riders on the platform.[3]

Workers have a productivity level on the Platform, z, which may differ between disabled and non-disabled workers, and based on access to disability-focused AI tools. The share of total work time allocated to platform work is denoted by h, while $1-h$ represents the share of time spent on other work pursuits. Output on the platform is given by zh, and workers are compensated at a piece rate w based on their output. The return to work off the Platform is given by $\frac{1}{\theta} r \left(1-h\right)^{\theta}$, where r measures the value of one's outside option and $0 < \theta < 1$.

Therefore, worker utility is given by:

$$U = w \cdot zh + \frac{1}{\theta} r \left(1-h\right)^{\theta} \tag{1}$$

In our empirical analysis, we measure productivity broadly, incorporating speed at different points, customer reviews, and average total time per order.

Workers make a decision about the share of their work time to spend on the Platform, with our first order condition:

$$\frac{\partial U}{\partial h} = wz - r\left(1-h\right)^{\theta-1} = 0 \tag{2}$$

From equation (2), we can write the hours worked as:

$$h = 1 - r^{\beta} \left(wz\right)^{-\beta} \tag{3}$$

The parameter $\beta \equiv \frac{1}{1-\theta}$ governs how responsive the worker's labor supply is to changes in productivity or the outside option. In other words, β captures how sharply the worker substitutes between platform and non-platform work. A *larger* β, corresponding to θ approaching 1, indicates that a small increase in the wage or productivity results in a *large* shift in platform work hours. Since we assume $0 < \theta < 1$, it follows that $\beta \in (1, \infty)$.

[3]We detail our institutional setting in Section 3 and present our ethnographic observations, including interviews and fieldwork, in the Appendix.

From equation (3), taking the partial derivative of h with respect to r and z, we can get:

$$\frac{\partial h}{\partial r} = -\beta r^{\beta-1} (wz)^{-\beta} < 0 \tag{4}$$

$$\frac{\partial h}{\partial z} = \beta r^{\beta} w (wz)^{-(1+\beta)} > 0 \tag{5}$$

Based on the literature and our ethnographic observations, we posit two main differences between disabled and non-disabled workers:

Assumption 1.

$$r_D < r_N$$

First, we assume that disabled workers have worse outside options than non-disabled workers (i.e., $r_D < r_N$). This is consistent with a large body of economic literature on the challenges and discrimination faced by disabled workers.

Assumption 2.

$$z_D < z_N$$

Second, in the absence of helpful technologies or other accommodations, we posit that disabled workers may have lower productivity than non-disabled workers.

There are three main ways to think about behavior on the Platform: efficiency (z), labor supply (h), and total output (zh). Based on the implications of the partial derivatives of h with respect to the outside option r and productivity z in equations (4) and (5), we derive several predictions:

Claim 1. Before the new AI tool, differences in working time between disabled and non-disabled workers are ambiguous.

The nature of this ambiguity depends on the relative differences in r (outside options) and z (efficiency). If the difference in outside options is large relative to differences in efficiency, disabled workers will work more hours on the Platform than non-disabled workers. Conversely, if the difference in outside options is smaller than the difference in efficiency, disabled workers will work fewer hours.

Claim 2. Before the new AI tool, the average differences in total output and profits between disabled and non-disabled workers are ambiguous.

If the difference in outside options is large, then disabled workers may have higher output and profit than non-disabled workers.

Claim 3. Suppose a new technology increases the efficiency of disabled workers. This will also increase their working time, total output, and profits on the platform, thereby closing the gaps between disabled and non-disabled workers.

Firm profits from a worker on the Platform are given by

$$\pi = -c \cdot \mathbb{1}_{\{\text{hired}\}} + (1 - w) \cdot zh - k \cdot \text{Bad Rating}, \tag{6}$$

where $\mathbb{1}_{\{\text{hired}\}}$ is an indicator function that equals 1 if the worker is newly hired and 0 otherwise. The term $(1 - w)\, zh$ represents the portion of total output (zh) that remains with the firm after paying the worker, whose compensation is $w\, zh$. The term $-c\, \mathbb{1}_{\{\text{hired}\}}$ captures the on-boarding or hiring cost, which is incurred only when a new worker is hired.[4] The term $k \cdot$ Bad Ratings represents the cost of bad ratings, where k is the cost to the Platform per bad rating and Bad Ratings is the number of bad ratings a driver receives.

3 Institutional Context and Data

We study a large Chinese food-delivery platform (the "Platform"), one of the two leading companies in the country's massive online food-delivery market. This Platform operates similarly to services like Uber Eats or DoorDash but on a much larger scale. Together, the two main platforms in China engage approximately 10 million delivery workers, making this a labor force larger than that of some entire countries. On our Platform, drivers complete hundreds of millions of deliveries per month.

Delivery workers on the Platform use motorbikes or e-bikes to transport food, groceries, and other items from restaurants, convenience stores, and pharmacies to customers. The service is available nationwide, with a higher concentration of workers in urban and eastern regions where food delivery is most prevalent.

Hiring policy. A key feature of the Platform is its open hiring policy: anyone who meets the basic requirements——such as minimum age, possession of a valid driver's license, and no disqualifying criminal record——can register and begin working. Importantly, there are no separate or stricter hiring standards for disabled applicants. Thus, disabled workers in our study are not subject to a different hiring standard relative to non-disabled workers.

Work models. Workers on the Platform can choose between two work models: a team-based model, where workers are assigned to stations and may receive a structured wage scheme (with unit prices following a step function based on the number of monthly completed

[4] We can allow the on-boarding costs for disabled and non-disabled workers to differ.

orders), or gig work, where workers operate independently and are paid per delivery. The two groups also differ in task assignment. Team workers receive pickup assignments from the platform's algorithm, while gig workers claim pickups on a first-come, first-served basis through the app. For team workers, order assignment is prioritized based on the number of past completed orders, but disability status is not considered. Approximately two-thirds of the workers are gig workers, as shown in Table 1. Both worker types are included in our analysis, and we control for these differences in our empirical models.

3.1 Assignment of Workers to Deliveries.

Order assignment is determined by three key metrics: the service score, rider level, and order acceptance rate.

Assignment Score = Service Score + Rider Level + Acceptance Rate

The service score primarily depends on customer reviews (particularly negative reviews) and instances of late deliveries. The Rider Level is calculated based on tenure and the cumulative number of completed orders. Refusing system-assigned orders directly reduces a rider's Acceptance Rate. Workers report that this factor has the greatest impact on assignment score, so they usually do not reject system-assigned tasks. Notably, a rider's disability status does not affect their order assignment score.

The order assignment score influences two key operational aspects. First, order assignment priority is primarily determined by the assignment score, but the system also takes logistical efficiency into account. Orders are assigned in a way that aligns with a rider's current route, optimizing for route consistency and operational effectiveness. Second, the assignment score affects the maximum number of concurrent orders a rider can handle. The maximum number of concurrent orders starts at a baseline of 3, with an upper limit of 13 for gig workers and 20 for team workers. A higher assignment score results in a greater allowable maximum. Typically, experienced workers handle an average of six concurrent orders, while high-performing workers can manage even more, often reaching around ten concurrent orders.[5]

Critically, disability status is not used in the assignment of workers to deliveries. Prior to AI, customers got an in-app message informing them their delivery worker was disabled,

[5]Gig workers and team workers have distinct order allocation methods. Gig workers primarily rely on independently selecting orders from the order pool. While they do receive system-assigned orders, these are preferentially assigned to team workers under the same conditions. On average, 20% of gig workers' orders come from the system, while the remaining 80% are manually selected by the workers themselves. Gig workers operate across the entire city, often covering multiple commercial districts. In contrast, team workers almost entirely depend on system-assigned orders. The typical assignment range for team workers is within six kilometers of their designated station.

but drivers still found it hard to communicate with customers.

Data

Our dataset is a proprietary administrative panel provided by the Platform, covering a three-year period from 2021 to 2023. The unit of observation is the worker-week—that is, for each worker in each week, we have detailed data on their activity and performance. The sample includes:

- All of the $4,284$ hearing-impaired workers who registered and worked on the platform between 2021 and 2023.
- $40,000$ non-disabled workers, randomly selected from the pool of all workers who registered in the same period, to serve as a comparison group.
- In total, we have hundreds of thousands of worker-week observations.

This dataset allows us to track each worker from the start of their tenure on the platform and analyze how their performance evolves over time. Importantly, by focusing on newly registered workers, we can compare disabled and non-disabled workers from their point of entry, ensuring a fair and meaningful comparison.

The AI tool was rolled out suddenly in April 2022 to all hearing-impaired workers. We have one year of pre-AI data to compare the performance of disabled and non-disabled workers, and we can track workers for more than a year to study the AI tool's impact.

We examine a wide range of outcome variables capturing different aspects of worker performance, categorized into three main groups:

1. Total Output (zh in our model): Measures the number of orders completed per week and the coverage of serviced areas, including the total number of unique restaurants and customers served.
2. Efficiency (z in our model): Captures service quality and the average time spent per delivery, further broken down into time spent at each stage of the work process.
3. Labor Supply and Retention (h in our model): Includes total hours spent online per working day, the number of days worked per week, and the worker's retention rate.

We summarize all outcome variables below:

There are three steps in the work process, and at each step, the worker must perform an action on the app, which provides a clear time stamp for when each step starts and finishes.

Key Outcome Variables by Category

1. Total Output (zh)	**2. Efficiency** (z)	**3. Labor Supply** (h)
Total Orders Delivered	Order → Worker	Total Hours Online
Total Orders in Peak Hours	Worker → Store	Number of Working Days
Number of Restaurants Served	Store → Delivery	Working Days During Bad Weather
Number of Customers Served	Ratings (Number of Bad Reviews)	Survival Rate
	Overtime Rate (Number of Overtime Orders)	

We do not have this information at the order level, but we do have the average time spent per order at the worker-week level.

The three steps are as follows: First, once an order is initiated by the customer, it is assigned to a worker (Order → Worker). Second, once the worker receives the order, they pick it up from the restaurant (Worker → Store). The waiting time for food preparation is included in this step, which concludes when the worker actually receives the food from the restaurant. Lastly, the worker delivers the order to the customer (Store → Delivery).

Additionally, our dataset includes various control variables that account for worker characteristics and job conditions. These variables include worker demographics (e.g., age, gender), city of work, worker type (team-based vs. gig worker), registration cohort (year of entry), and order attributes (e.g., average customer-to-restaurant distance). By incorporating these controls, our empirical analysis ensures that comparisons between disabled and non-disabled workers are made under similar conditions.

Table 1 presents the summary statistics for the sample of 43,605 food-delivery workers. Demographically, the disabled workers in our sample are slightly older (mean age 32.82 vs. 30.29) and more likely to be male (92.55% vs. 90.82%) compared to their non-disabled counterparts; these differences are statistically significant at the 1% level. However, regarding the nature of their employment, we find no statistically significant difference in the registration type between the two groups. Approximately 35-36% of workers in both groups are registered as "Team Workers", while the remaining 64% are "Gig Workers", suggesting that disability status does not dictate the contractual arrangement with the platform. Panel C details the distribution of disability severity among the hearing-impaired workers, showing a high concentration in the most severe category (Level 1), which accounts for 53.57% of the disabled subsample.[6]

[6]In contrast to our sample where over 53% of disabled workers fall into the most severe category (Level 1), the general population of individuals with hearing impairments exhibits a much lower prevalence of severe disability. National data indicates that the prevalence rate for "Extremely Severe" (Level 1) hearing impairment is 15.4%. This discrepancy suggests that our sample is skewed toward those with the most profound hearing loss.

Table 2 provides a detailed summary of key performance indicators for food-delivery workers, categorized into Total Output, Efficiency, and Labor Supply on the Platform.

On average, workers complete about 130 deliveries per week, with a median of 89. The standard deviation is quite large (about 140), indicating considerable variation in total deliveries. Peak-hour orders, which occur during high-demand times, account for about 77 deliveries per worker per week, suggesting that a substantial share of total orders happens during peak times. Team-based workers complete significantly more deliveries on average (103) compared to individual workers (25), reflecting the full-time nature of team-based work.

The time spent in different stages of the delivery process follows some notable patterns. The order-to-worker assignment time is relatively short, averaging 3.4 minutes, though the median is much lower at 2.4 minutes, indicating that most deliveries are assigned and accepted quickly, but a few take much longer. The worker-to-store travel time averages 11.6 minutes, with a median of 11.4 minutes, suggesting fairly consistent travel times. The store-to-delivery time is slightly longer at around 14 minutes, with a standard deviation of 5.6, implying variability due to delivery distances and potential communication delays with customers. Customer satisfaction, measured as bad ratings per 10,000 orders, averages 9.6, though the median is 0, indicating that most workers receive very few negative reviews, but a subset accumulates significantly more complaints.

Workers spend an average of about 1,520 minutes online per week, with a median of around 590 minutes. The high standard deviation (about 1,800) indicates wide variation in work intensity, with some workers putting in significantly more hours than others. The median number of workdays per week is 6, though the average is lower at 4.5, suggesting that while many workers work full weeks, others participate part-time or intermittently. The median tenure on the platform is 3 weeks, but the average is 11 weeks, with a high standard deviation, indicating that while many workers leave quickly, a significant portion remain active for longer durations.

In summary, the data reveal substantial variation in work intensity among workers—while some deliver a high volume of orders, many operate on a more part-time basis, as evidenced by the wide distribution in orders completed and hours worked. Team-based workers are notably more productive, handling nearly four times as many orders as gig workers, suggesting structural differences in work allocation and efficiency. Finally, while retention is high for some workers, a significant portion of workers do not stay on the platform for long periods.

Ethnographic Observations. In addition to analyzing quantitative data, our research team conducts qualitative fieldwork to better understand the experiences of disabled workers, building on this practice in past RCTs like Blader et al. (2020) and Friebel et al. (2025).

This included:

1. Interviews with 10 hearing-impaired workers to learn about their daily routines, challenges, and communication methods.

2. On-the-ground field observations were conducted, during which researchers accompanied disabled workers and delivered orders full-time alongside them. In total, the researchers completed slightly more than 100 deliveries over multiple days. The purpose of these observations was to gain firsthand insight into the work environment of disabled workers, including their delivery process and customer interactions.

These qualitative insights helped us interpret patterns in the data and identify specific barriers faced by hearing-impaired workers—particularly in areas such as customer communication, navigation, and service speed. They also provided context for understanding labor supply differences, as many hearing-impaired workers expressed a strong attachment to their jobs due to limited alternative employment opportunities. Details of the interviews, photos, and fieldwork documents are provided in the Online Appendix C.

Together, the combination of high-frequency administrative data and direct ethnographic observation provides a rich and comprehensive view of the productivity, efficiency, and retention of disabled workers in the gig economy. This dataset and research approach form the foundation for our empirical analysis.

4 Pre-AI Differences in Productivity, Compensation, Retention, and Profitability

Our analysis proceeds in two stages. First, in Section 4, we measure differences in performance, retention, compensation, and profitability between disabled and non-disabled workers prior to the introduction of the AI tool. Second, we evaluate the impact of the AI tool on the outcomes of disabled workers (versus a comparison group of non-disabled workers) using a difference-in-differences approach.

4.1 Empirical Strategy and Econometric Specifications

Using data from before the technological intervention in April 2022, we first compare the job performance of disabled (hearing-impaired) workers to that of their non-disabled counterparts. To estimate the average differences between these groups, we employ a regression framework. We regress each outcome of interest on a disability indicator and a set of controls.

The baseline specification is:

$$Y_{it} = \alpha + \beta\, Disability_i \;+\; \mathbf{X}'_{it}\gamma \;+\; \mu_t \;+\; \chi_i \;+\; \varepsilon_{it}, \tag{7}$$

where worker i works during week t, Y_{it} represents the outcome for worker i in week t. The term $Disability_i$ is a dummy variable indicating whether worker i is hearing-impaired (disabled) or not. To adjust for differences in the nature of work performed, we include a vector of time-varying controls, $\mathbf{X}_{it}$, such as the average customer-to-restaurant distance of orders that week. Additionally, we incorporate a set of worker characteristics, χ_i, including their city of work, team worker status, age, gender, and registration year. The term μ_t represents the time fixed effect. We apply this analysis across multiple outcome categories, including efficiency metrics, output metrics, and labor supply metrics.

Next, we investigate the dynamics of worker productivity by estimating productivity-tenure curves separately for disabled and non-disabled workers. We employ a regression approach in which each parameter β_t captures the average productivity level of workers at tenure t months on the platform. Specifically, we estimate the following equation:

$$y_{it} = \sum_t \beta_t \times \mathbb{1}[\text{Tenure}_i = t] + \mathbf{X}'_{it}\delta + \epsilon_{it}, \tag{8}$$

where y_{it} denotes the productivity outcome for worker i in week t. The key explanatory variables are a set of tenure indicators ($\mathbb{1}[Tenure = t]$), allowing the estimation of productivity at each specific tenure level. We include a vector of time-varying controls ($\mathbf{X}_{it}$), such as the average customer-to-restaurant distance, to account for systematic differences in order characteristics that could influence worker productivity. Additionally, year-month fixed effects and city fixed effects are incorporated into the regression model to control for common temporal shocks, seasonal patterns, and location-specific factors that may influence productivity independently of worker tenure.

Throughout the paper, standard errors are clustered at the worker level, as disability status varies by worker.

4.2 Results

Table A.2 summarizes regression results for key outcome measures prior to the introduction of the new technology, where each coefficient represents the estimated difference associated with the Disability indicator (disabled = 1) from equation (7), controlling for various factors described earlier. For ease of exposition, these performance differences are illustrated in Figure 1. The bar plots depict effect sizes standardized by sample standard deviations,

providing a clear visual comparison between disabled and non-disabled workers.

Total Output. Disabled workers deliver more orders per week on average than non-disabled workers. Specifically, being a disabled worker is associated with completing approximately 3 additional orders per week (2% more than non-disabled workers). They also complete about 2 more peak-hour orders per week (during high-demand periods). These differences are substantial, given that the median number of weekly orders in the sample is only a few dozen. The positive output differential is even larger within the subset of Team deliveries: disabled team workers handle approximately 10 more orders per week (4.7% more) than their non-disabled counterparts. Conversely, in the Gig (independent) segment, disabled workers complete about 5 fewer orders per week (9.6% less) than non-disabled gig workers.

This likely reflects a sorting or selection effect: the Team format provides more stable order assignments, and team workers can seek help from station managers, whereas purely gig-based workers are more independent and rely entirely on themselves. Nonetheless, aggregating across types, the total orders delivered per week is higher for disabled workers. Furthermore, disabled workers tend to serve a broader range of customers and merchants: they serve on average 4 more restaurants and 7 more unique customers (per week) than comparable non-disabled workers. These metrics reinforce the idea that disabled individuals are providing a high volume of service on the platform. The finding that disabled workers have higher total output aligns with our expectation that their greatly increased work time (to be discussed shortly) more than compensates for any per-order speed disadvantage. It also directly answers part of our first research question: in terms of sheer output, the firm gains somewhat more from an average disabled worker than from a non-disabled worker in the short run.

Efficiency (Speed and Quality). Despite their higher output, disabled workers do exhibit some efficiency shortfalls on a per-order basis. In our data, disabled status is associated with slower delivery times in each segment of the order fulfillment process. For the stage from order initiation by the consumer to order assignment to the worker, disabled workers take about 0.16 minutes longer per order on average (approximately 5% slower than non-disabled workers). For the stage from confirming the order assignment to departure—which includes traveling to the restaurant, communicating with restaurant staff, and picking up the order—they are about 0.21 minutes slower (1.7% slower). The largest gap occurs in the final stage, from leaving the restaurant to delivering the order to the customer: disabled workers take about 1 minute longer (approximately 5.6% more) on average.

All these differences are statistically significant. While being less than a minute slower on average may seem minor for a single delivery, across many deliveries, the time adds

up. This pattern aligns with insights from our interviews, which indicate that a major hurdle is communicating with customers upon arrival. Our fieldwork reveals that, in most cases, disabled workers are just as efficient as non-disabled workers. However, there are occasional situations where workers need to communicate extensively with customers, leading to significant delays.[7] In China, delivery workers are generally expected to call the customer upon arrival—for instance, to coordinate the handoff or confirm the delivery if the food is left in a delivery cabinet or at the front desk.

Prior to 2022, hearing-impaired workers who could not speak had to rely on app messaging or make a call, wait, and hope the customer noticed the message. This often led to delays – which likely explains why the Store → Delivery time segment (covering the travel to customer and the handoff) showed the largest delay. Additionally, disabled workers might drive slightly more cautiously or slower due to their hearing impairment (for safety), contributing to these time gaps. We find that disabled workers have a higher chance of delivery delays. They deliver 34 more overtime orders per 10,000 orders compared to non-disabled workers. We also find that disabled workers receive lower customer ratings on average. Specifically, they receive about 5 more "bad" reviews per 10,000 orders delivered (33% more) compared to non-disabled workers.

In absolute terms, customer complaints are rare for both groups (the baseline bad review rate is low), but this difference is meaningful – it suggests that hearing-impaired workers were more likely to have instances of customer dissatisfaction. Customers might become frustrated if they cannot communicate easily with the worker (for example, if a worker cannot call to find them), or may misinterpret a lack of spoken communication as rudeness. Indeed, some hearing-impaired workers in interviews recounted instances where customers didn't realize the worker was hearing-impaired and left poor feedback about "not calling" or "not speaking clearly."

This gap in service quality is an important finding because it hints at a potential long-run risk: if customers are less satisfied, they might order less or the platform's reputation could be affected. The platform noted this concern, which is one reason they invested in the calling technology. In summary, before the intervention, disabled workers were slightly less efficient per order and had lower customer service ratings than their peers – confirming that $z_D < z_N$ in our model terms. However, these efficiency gaps did not stop disabled workers

[7]For example, during one delivery, when we were halfway to delivering the food, we received a phone call from the restaurant owner. She said she had forgotten to add fresh tomato to the package, as the customer had requested. She asked if we could return to the restaurant to pick it up. We had to explain that we were already halfway and couldn't go back without delaying the order. At the same time, we also needed to ask the restaurant owner to call the customer directly and explain the situation in case the customer blamed us and left a bad review.

from contributing strongly to the platform, thanks to their work effort, as we discuss next.

Labor Supply. Perhaps the most striking differences are in labor supply. Disabled workers tend to provide significantly more labor on the platform. Our regression results show that, controlling for other factors, disabled workers spend about 25 more minutes online per week (3.66% more). This finding is consistent with the notion that the disabled workers have fewer alternative opportunities and hence are willing to work more hours (i.e., their effective r is lower, driving a higher h^*). We also find disabled individuals work slightly more days per week: about 0.3 additional work days on average (5.18% more). It is not uncommon for hearing-impaired workers to work 6 or even 7 days a week, whereas some non-disabled workers might treat this as a secondary job and work only weekends or a few days. Furthermore, disabled workers are more likely to work during unpleasant conditions, with about 0.05 more days in bad weather (rain, etc.) per week than others.

This suggests a strong commitment – even when conditions are unfavorable, they tend to keep working (possibly because they really need the earnings). Finally, and importantly, disabled workers have substantially higher retention rates on the platform. We measure a "survival" metric indicating whether a new hire is still active after a certain period. The regression coefficient of 0.0628 can be interpreted as an approximate 12 percentage-point higher probability of remaining active at a given point (we observe them up to the end of 2023).

In practical terms, hearing-impaired workers are much less likely to quit. Many non-disabled workers treat the job as temporary—turnover is high—but disabled workers often remain on the platform longer, perhaps due to the scarcity of stable job opportunities for them. We confirmed this pattern by plotting Kaplan-Meier survival curves (Panel (a) of Figure 8): the curve for disabled workers lies significantly above that of non-disabled workers, indicating that a higher fraction of disabled workers continue working months after joining. We further explore onboarding and retention in the next subsection.

Onboarding and Retention. Figure 8 illustrates survival curves based on weeks worked. During the onboarding phase, defined as the period after individuals register as potential workers but before they deliver any orders, the exit rate for disabled workers is notably higher compared to their non-disabled counterparts. Specifically, 22% of non-disabled registrants never begin delivering orders, whereas this percentage increases for disabled workers (38%), indicating a significant onboarding challenge among disabled individuals.

Qualitative insights from interviews identify key barriers faced by disabled workers during onboarding, primarily inadequate support and technological hurdles. Disabled workers frequently report difficulties such as unfamiliarity with restaurant locations, challenges in navigating apartment complexes, and significant mental stress related to potential commu-

nication issues with restaurant staff or customers in case of disputes.

However, Figure 8 also reveals a two-stage retention dynamic. After the initial challenging months, both disabled and non-disabled workers experience decreased attrition rates. Notably, disabled workers who surpass the initial onboarding period show a marked stabilization in employment, typically around eight months. This stabilization indicates that once initial barriers are overcome, disabled workers tend to achieve comparable or even superior long-term retention rates compared to non-disabled workers.

Interview data further corroborate this finding, suggesting that long-term retention among disabled workers is greatly facilitated by increased familiarity with their work environment, including neighborhoods, frequently visited restaurants, and common delivery destinations such as condominium complexes. Additionally, team workers emphasize the pivotal role of managerial support (from station managers) and peer support (from other disabled workers in the same team) in fostering job satisfaction and loyalty.

These findings underline a two-stage retention process: a critical initial phase marked by large attrition for disabled workers due to onboarding difficulties, followed by a stable employment period characterized by robust retention. Improving onboarding practices—such as providing targeted accommodations, promoting social inclusion, and offering strong managerial support from day one—could effectively reduce early attrition among disabled workers. Addressing these early challenges not only closes the initial retention gap but also harnesses the high potential for loyalty and long-term attachment that disabled workers frequently demonstrate once integrated into their roles.

Worker Compensation. Panel A of Table A.1 reports results comparing disabled and non-disabled workers in terms of compensation and profit. We use inverse hyperbolic sine to account for the fact that these variables can be zero or negative, though we obtain quite similar results for income and hourly wage in log terms. As seen in column 1, disabled workers have 4% higher monthly income relative to non-disabled workers, reflecting that they work more hours on the Platform. However, when we focus on hourly pay, we find that disabled workers earn 8% less than non-disabled workers, reflecting primarily that they are slower and are also more likely to financially penalized for bad ratings. The disability wage gap in our stetting of 8% is broadly consistent with recent work by Forth and Jones (2025). They estimate a within-firm disability pay gap of 6-7% using census data on workers in many firms in England and Wales.

4.3 Profits

We use the estimated performance differences to estimate the profit disparity between disabled and non-disabled workers. Disabled workers tend to stay longer and complete more orders each month, but are also more likely to receive negative reviews.

Using our definition of the profit function in equation (6), the profit difference between disabled (D) and non-disabled (N) workers can be formally expressed as:

$$\Delta\pi = \pi_0 \cdot E(Y^D - Y^N) - c \cdot \Delta Pr(\text{hired} = 1) - k \cdot E(\text{Bad Ratings}^D - \text{Bad Ratings}^N), \quad (9)$$

where Y^D and Y^N represent the number of orders completed by disabled and non-disabled workers, respectively; π_0 represents the average profits per order received by the Platform; c represents the cost of turnover, which is the cost of on-boarding a new worker; $Pr(\text{hired} = 1)$ captures the differential probability of a worker being hired; k represents the cost of a bad rating; and Bad Ratings^D and Bad Ratings^N represent the number of bad ratings received by disabled and non-disabled workers, respectively.

According to Table A.2, disabled workers complete an average of 2.3 more orders per week than their non-disabled counterparts. The Platform earns a profit of about 7-8 CNY per order. Therefore, the Platform earns about 16-18 CNY ($2 USD) per worker-week from disabled workers doing more orders.

Furthermore, disabled workers exhibit a higher retention rate. Table A.2 shows that disabled workers have a 6.3 percentage point higher probability of continued employment (survival probability) per month. We assume an onboarding cost of 100 CNY. Thus, disabled workers yield an incremental profit of 6 CNY relative to non-disabled workers due to higher retention.

Finally, we account for bad ratings. Disabled workers receive 2 more bad ratings per 10,000 orders compared to non-disabled workers, which works out to about 0.02 more bad ratings per worker-week. Based on our interactions with the Platform, a reasonable cost to the Platform of a bad rating is 100 CNY, a value which is designed to account for the full cost of such a rating, including effects on customer retention and reputational spillovers. Thus, disabled workers cost the firm about 2 CNY relative to non-disabled workers.

Column 3 of Panel A of Table A.1 estimates that disabled workers earn 3% *higher* profits for the firm than non-disabled workers. This estimate narrowly misses statistical significance, so we cannot reject that disabled and non-disabled workers earn similar profits for the firm. However, we can strongly reject that disabled workers are less profitable. With 95% confidence, we can reject that disabled workers are more than 1% less profitable than non-disabled workers in the pre-AI period.

5 Evaluating the AI Intervention

5.1 Origins of the AI tool

The "Intelligent Outbound Call" is a feature on the app specifically designed to empower disabled delivery workers, particularly those with hearing and speech impairments, by enhancing their ability to communicate effectively with customers during delivery processes. Before this AI technology, disabled workers were restricted to communicating solely through text-based messages within the delivery platform's app. This posed significant challenges, as typed messages such as "Can you open the door for me?" or "I cannot find the place, please allow me an extra five minutes" often went unnoticed or unanswered by customers, resulting in delays, misunderstandings, and added stress for the workers.

In late April 2022, the platform introduced the "Intelligent Outbound Call" (IOC) feature, which allows disabled workers to initiate automated voice calls to customers directly through the app. The tool was developed in-house at the Platform by software engineers during off time. In its messaging to the public and employees, the Platform emphasizes corporate social responsibility, and part of the goal of the AI tool was to help disabled workers.

Under the IOC feature, each call begins with an introductory message clearly stating that the caller is a disabled worker using an automated system due to communication barriers, followed by a specific message conveying the worker's issue or request (for example, "Your food will be placed in a locker"). After selecting a prompt, the app initiates an automated voice call to the customer.

This function is based on Text-to-speech (TTS) technology, which converts written text into natural-sounding speech and is widely used in voice assistants, audiobooks, and accessibility services. While traditional methods could perform basic functions, pre-AI generated speech often lacks naturalness and emotion. AI has brought significant breakthroughs to this area. Deep learning-based models (such as WaveNet and Tacotron) use neural networks to directly synthesize speech waveforms, significantly improving fluency and making prosody, intonation, and pauses more human-like. AI also enables rapid multilingual adaptation and diverse style simulation, capable of replicating specific voice characteristics and expressing different emotions.

The system provides workers with a selection of default prompts (Panel (a) of Figure D.1) or the option to create and customize prompts tailored to specific situations. In Online Appendix D, we include a screenshot showing commonly used customized prompts by disabled workers (Panel (b) of Figure D.1). These prompts typically fall into one of the following categories: (1) difficulty locating the customer, such as when security denies the worker entry

or when a buzzer code is needed; (2) notifying the customer that the order has been placed at a specific location, such as a food cabinet; or (3) politely requesting a positive review.

The advantages of the "Intelligent Outbound Call" technology include: first, enhanced customer responsiveness, as customers are significantly more responsive to phone calls compared to app notifications, ensuring prompt attention to critical communications; second, improved efficiency, since disabled workers can quickly address common issues through preset prompts without repeatedly typing messages, saving valuable time especially during peak delivery hours or adverse weather conditions; and third, personalization and flexibility, as workers can customize messages to clearly and effectively communicate unique situations.

5.2 Overall Effects

To assess the impact of the Intelligent Outbound Calling feature, we employ a difference-in-differences (DiD) research design. The basic idea is to compare the changes in outcomes for the treated group (disabled workers, who benefited from the new feature) before vs. after the intervention, relative to the changes experienced by a control group (non-disabled workers, who were not directly affected by the feature).

Let *Post* be an indicator variable that equals 1 in the period after the introduction of the new technology (Intelligent Outbound Calls, late April 2022 onwards) and 0 before. Our DiD regression can be specified as:

$$Y_{it} = \gamma \, \mathbb{1}[Disabled_i = 1] \times Post_t \; + \; \lambda_i \; + \; \xi_t \; + \; \varepsilon_{it}. \tag{10}$$

The interaction term $Disabled_i \times Post_t$ captures the DiD estimate of interest, and γ measures the differential impact of the technology on disabled workers relative to their non-disabled counterparts.

We include worker fixed effects λ_i and time fixed effects ξ_t. Worker fixed effects control for any time-invariant differences across individuals, while time fixed effects (weekly or monthly) capture trends affecting all workers. We estimate equation (10) using panel data at the worker-week level.

Figure 3 summarizes the treatment effects of the technology on all our outcomes. The bar plots depict effect sizes standardized by sample standard deviations, providing a clear visual comparison between disabled and non-disabled workers. The original regression results are reported in Table A.3.

Total Output. The introduction of Intelligent Outbound Calling (IOC) led to a significant increase in the number of orders completed by disabled workers. Treated workers

(hearing-impaired) delivered about 6 more orders per week because of the technology, representing a 4.7% increase compared to the average number of weekly orders for disabled workers prior to the intervention. This is an additional boost on top of their existing output advantage – effectively, the gap in weekly orders between disabled and non-disabled widened further post-intervention, in favor of disabled workers. The increase is particularly notable in the gig segment, where disabled workers delivered 9 more orders per week than they otherwise would have—an 18% increase compared to their pre-technology intervention average.

Interestingly, recall that pre-AI, disabled gig workers were delivering fewer orders than non-disabled gig workers (a -5 gap). After IOC, not only did they catch up, they surpassed non-disabled gig workers by a considerable margin. In the team segment, there was a positive but smaller and only marginally significant effect of an additional 3 orders. This suggests that IOC particularly helped those hearing-impaired workers who were working in the more independent gig mode – possibly because team settings already had some support or workarounds in place, whereas gig workers really benefited from the feature. In terms of other output metrics, disabled workers post-IOC served more restaurants (+4) and more customers (+6) than they would have without the technology. All these gains are statistically significant.

The bottom line is that the technology allowed disabled workers to handle an even larger volume of orders. This could be due to multiple reasons: faster turnover per order (if each order takes less time, they can fit more in a day), higher customer retention (if customers complain less, workers stay active and get more orders), or even the platform's algorithm giving them more orders if their metrics improved. Our analysis cannot pinpoint the exact mechanism, but the net effect is clearly a boost in output. For the platform, this means higher revenue per disabled worker after the intervention.

Efficiency and Service Quality. The IOC feature was aimed at improving communication, so we expect to see improvements in the efficiency metrics, particularly those related to the delivery handoff. Indeed, the data show that after IOC, the time spent in the final stage (store to delivery) by disabled workers decreased significantly relative to non-disabled: a reduction of 0.3 minutes per order on average. This indicates a substantial closing of the gap – recall the pre-IOC disabled lag in this stage was 0.82 minutes, so roughly one-third of that gap was eliminated immediately after introducing the call feature. Interestingly, for the earlier steps (Order $\rightarrow$ Worker dispatch time, and Worker $\rightarrow$ Store travel time), the coefficients are small and not statistically significant. This suggests those steps in the work process were not significantly affected by the calling feature, as expected, since those stages don't involve customer interaction). Thus, IOC specifically tackled the delay at the delivery

stage, which was indeed the main pain point.

Crucially, the number of order delay events decreased dramatically—by 10 per 10,000 orders. Note that the raw difference between disabled workers and non-disabled workers is 34 orders out of 10,000 orders. A reduction of 10 overtime orders represents approximately a 30% decrease in overtime orders. As a result, customer satisfaction for disabled workers improved markedly. The number of bad reviews per 10,000 orders for disabled workers dropped by about 2 relative to the control trend. This is a significant improvement, given that before IOC, disabled workers had around 5 more bad reviews per 10,000 orders than non-disabled workers. This reduction means that the gap in bad reviews between disabled and non-disabled workers shrank significantly. Customers presumably had fewer reasons to complain once they started receiving the normal phone call alerts (albeit automated) from disabled workers. The communication friction was alleviated, leading to smoother delivery completions and happier customers.[8] Qualitatively, our interviews with workers after IOC indicated they felt more confident and had fewer incidents of misunderstanding. Some customers even left positive feedback, appreciating the effort when they realized that the worker was hearing-impaired and was using an assistive calling system.

Labor Supply and Retention Post-Intervention. Our diff-in-diff results show no significant change in total hours online for disabled workers attributable to the IOC introduction. The coefficient on online duration is negative (roughly -1.4 hours per week if averaged, but with a very large standard error) and not statistically different from zero. Thus, we cannot reject that disabled workers kept their heavy work hours roughly the same after getting the new tool. They did slightly increase the number of days worked per week, by about 0.14 days. This could mean a few more disabled workers decided to work a 7th day occasionally when they might not have before, perhaps feeling more empowered. Notably, however, disabled workers significantly increased their bad-weather working days by 0.05 (a 7.7% increase compared to the pre-IOC mean), indicating greater willingness to work in tough conditions. Typically, during bad weather (heavy rain or heavy snow), there is a higher likelihood of food preparation delays, increased order volume, and more unexpected situations, requiring workers to communicate more frequently with customers.

Importantly, the retention of disabled workers improved further. The survival probability of disabled workers rose by an additional 6.2 percentage points post-IOC relative to non-disabled. This is a striking effect on top of an already higher baseline retention. It implies that the technology not only helped in performing the job but may have increased job satisfaction or viability for disabled workers, leading them to stay longer. In essence, IOC

[8] An alternative reason for the improvement in customer satisfaction may stem from the fact that customers notice the worker is disabled and are more sympathetic when giving ratings.

made the platform an even more attractive and sustainable workplace for disabled workers, which is a positive outcome both socially and for the firm (as it reduces hiring and training costs through lower turnover).

Monthly Income, Wages, and Platform Profit. Panel B of Table A.1 analyzes the impact of AI on disabled worker income, wages, and profitability. The AI tool increases worker income by 11%, reflecting the tool makes disabled workers work faster, have fewer bad ratings, and work more hours. The hourly wage effect increases by 5%, suggesting that roughly half of the effect on income is due to disabled workers working more hours. Given the pre-AI disabled pay gap of 8%, AI closes over half of the disabled pay gap. We are not aware of other interventions or technologies with comparable effects on disability wage gaps.[9] Column 3 of Panel B of Table A.1 increases the profitability of disabled workers by 11%, which is a very sizable and statistically significant increase.

In sum, the Intelligent Outbound Calling feature had a positive and significant impact on the productivity, efficiency, and retention of disabled delivery workers. It effectively boosted the productivity of disabled workers, substantially closing the performance gap between disabled and non-disabled, and in some aspects even giving disabled workers an edge.

Costs and Benefits of the AI Tool. The AI tool increases the profits per disabled worker by 11%, but how does this compare against the costs of the AI tool? The tool was developed in-house by Platform engineers, so there is no purchase or subscription cost to the tool. Assuming reasonable values for the time cost of the engineers who made it, the private benefits to the Platform are much larger than the costs.

5.3 Threats to Identification and Robustness Checks

Excluding the transition period. Like many new technologies, usage of the new AI tool was not instantaneous at 100% usage. Rather, according to the Platform, there was a transition period of a few months where disabled drivers learned about the new AI tool and how to best use it. This is supported by the event study graphs, which often show more muted impacts on workplace outcomes in the first few months, but increasing over time. Thus, to estimate longer-run impacts of the AI tool, we re-did our event study, but excluding the first 4 months immediately following the policy change. The results, presented in Figure A.5, confirm that our main findings are robust to this exclusion. In fact, the estimated magnitudes are generally larger when focusing on the post-transition period. For

[9]Evaluating the increase in work from home due to covid, Bloom et al. (2025) estimate that disabled worker wages fell more in occupations involving more work from home relative to ones involving less work from home. This reflects that work from home increases the labor supply of disabled workers, so that wages would naturally decline.

instance, the reduction in bad ratings intensifies to 23.9%, while total weekly orders and weekly income see substantial increases of 12.8% and 13.1%, respectively. This suggests that our baseline estimates, which include the transition period, likely represent a conservative lower bound of the technology's true impact on worker productivity and welfare.

Differential attrition. A potential concern with our baseline estimates is that the observed effects could be driven by compositional changes in the workforce rather than the causal impact of the AI tool. For instance, if the introduction of the technology induced lower-performing workers to exit the platform (attrition) or attracted different types of new entrants, our main results might reflect these selection effects. To address this threat to identification, we conducted a robustness check restricting the sample solely to a balanced panel of "stayer" workers, i.e. those observed working positive hours both before and after the introduction of the AI tool. As shown in Figure A.6, the results from this restricted sample are quantitatively and qualitatively consistent with our baseline findings. We continue to observe significant improvements across key metrics: total weekly orders increased by 11.6%, weekly income rose by 11.8%, and bad ratings decreased by 17.8%. This stability in coefficients indicates that our main findings are not a result of differential attrition or changing workforce composition, but rather reflect genuine productivity gains for the workers who use the technology.

Hours sample is different. As noted in the Data section, granular data on worker hours (online duration) is only available for the latter portion of the pre-AI period. Consequently, our baseline estimates for hourly-based metrics (such as hourly wages and efficiency per service hour) rely on a slightly different time window than other aggregate outcomes. To ensure that this sample restriction does not introduce bias, we re-estimated all key workplace outcomes restricting the sample solely to observations where online duration data is available. The results, presented in Figure A.7, demonstrate strong consistency with our main findings. Even within this restricted window, the introduction of the AI tool drives significant improvements: total orders per week increase by 13.3%, weekly income rises by 13.6%, and time spent between restaurant pickup and customer delivery decreases (-2.3%). These estimates align closely with the full-sample results, confirming that the timing of the hours data availability does not skew our general conclusions regarding the technology's impact.

The role of late deliveries and bad ratings in load assignment. A potential alternative mechanism driving our results is that the algorithm might have begun assigning "easier" or shorter-distance routes to disabled workers once their performance metrics (such as lateness and ratings) improved. If true, the observed efficiency gains would be a function of easier tasks rather than better communication. To address this, we re-estimated our

main specification while flexibly controlling for the difficulty of tasks, approximated here by the distance of the order. As shown in Figure A.8, our results remain robust even after accounting for trip distance. The coefficients for total orders (+11.1%), weekly income (+11.3%), and the reduction in bad ratings (−17.9%) are virtually unchanged from the baseline estimates. This stability suggests that the productivity gains are not an artifact of the platform assigning shorter or easier trips to treated workers, but rather stem directly from the efficiency improvements enabled by the IOC technology.

5.4 Heterogeneous Treatment Effects

A central question regarding the deployment of AI in the workplace is whether treatment effects vary across different segments of the workforce. For instance, recent work by Brynjolfsson et al. (2025) suggests that generative AI tends to confer larger benefits to novice or lower-skilled workers. To investigate similar dynamics, we examine heterogeneous treatment effects (HTE) using both standard interaction models and machine learning approaches for sorted effects (Chernozhukov et al., 2018).

We focus our analysis on three key dimensions: pre-AI tenure, pre-AI productivity, and the severity of the worker's disability. We find no evidence of heterogeneous treatment effects based on pre-AI tenure (Panel A of Figure A.12) or pre-AI productivity (Panel B), indicating that the technology's effectiveness in boosting output and wages is not contingent on a driver's prior experience or baseline skill level.

However, we observe significant heterogeneity with respect to the severity of disability. Figure 9 plots the coefficients for workers with the most severe hearing impairment (Level 1, $>$ 90 dB HL) versus those with moderate-to-mild impairments (Levels 2–4). For productivity and service quality metrics, the benefits are most pronounced for the most severely impaired workers. Specifically, Level 1 workers experience larger increases in weekly orders ($p = 0.08$ for the difference) and income ($p = 0.08$), and a significantly steeper reduction in bad ratings ($p = 0.06$) compared to their moderate/mild counterparts. This suggests that the AI tool is most effective at removing communication frictions for those who face the steepest barriers.

Interestingly, the pattern reverses for attrition ($p = 0.03$). The reduction in platform exit rates is driven almost entirely by workers with moderate/mild impairments (Levels 2–4), whereas retention rates for the most severely disabled workers (Level 1) remain statistically unchanged. A plausible interpretation is that severely impaired workers may have fewer outside labor market options (and thus higher baseline retention) regardless of the tool, whereas the technology effectively tipped the scales for moderate/mild workers who might otherwise have exited.

6 Survey Evidence and Next Steps

Our administrative analysis provides novel evidence on the productivity and retention of hearing-impaired workers, and on how a targeted AI intervention can enhance their performance. Yet administrative data alone cannot capture the full range of outcomes that matter for worker welfare or long-run inclusion. In particular, we lack measures of how disabled workers experience their jobs, how AI tools affect stress and job satisfaction, and what outside opportunities they face in the broader labor market. To address these gaps, we are designing and implementing a complementary survey of hearing-impaired and non-disabled workers on the Platform.

The survey will collect information on key dimensions such as job satisfaction, stress and mental health, experiences of discrimination, attitudes toward AI and technology, and outside options. These modules are designed to illuminate both the mechanisms behind our administrative results and the broader consequences of AI for workers' well-being. By linking the survey to the platform's administrative records, we can combine rich performance data with direct worker reports, providing a uniquely comprehensive picture of how AI shapes inclusion at work.

This next phase is important to our project. While our findings show that a simple AI intervention can substantially improve firm-side outcomes, understanding whether such technologies also improve the welfare and long-run prospects of disabled workers requires going beyond administrative metrics. The survey will allow us to assess these worker-side outcomes, evaluate the broader external validity of our results, and provide a fuller understanding of how AI can foster inclusivity in labor markets.

7 Conclusion

We examine the productivity of hearing-impaired food-delivery workers in China and evaluate an AI-based technological intervention designed to boost their performance. Even without accommodations, disabled workers made strong contributions: although they faced modest efficiency challenges on a per-order basis, they compensated through greater labor supply—spending more time online and being significantly less likely to quit. This pattern is consistent with our model, in which disabled workers face worse outside options and therefore remain more attached to platform work.

The introduction of a targeted AI solution—the Intelligent Outbound Calling feature—substantially enhanced performance. Delivery speeds improved, particularly in the last-mile stage involving customer interaction, while complaint rates fell sharply and weekly

order completions rose. Retention, already high, increased further. These gains narrowed efficiency gaps between disabled and non-disabled workers and raised the overall value of disabled workers to the firm. The intervention also had a large effect on reducing the gap between disabled and non-disabled workers in hourly wage. The intervention was relatively simple and inexpensive, relying on advances in text-to-speech technology to make automated calls sound natural, yet it generated large productivity and inclusion benefits. Far from being a limitation, the simplicity of the tool underscores that even more "modest" AI-enabled interventions (i.e., simpler than generative AI tools like large language models) can meaningfully reduce barriers faced by disabled workers. Further work can examine the impact of more advanced AI tools in affecting the work lives of disabled workers.

External validity. Our evidence also speaks to broader implications and external validity. In this setting, the private returns to adoption exceeded the costs, suggesting that firms may have incentives to implement such technologies without subsidies when the disability is common and the per-worker gains are high. For rarer disabilities, or where productivity improvements are smaller, public support may be warranted given the substantial social benefits of keeping disabled individuals employed and reducing reliance on disability insurance. Our key result, that AI can help make work easier and more lucrative (both personally and for the firm) for disabled workers, also shows promise informally in other contexts. Consider Elon Musk's Neuralink, which is manufacturing implantable brain computing devices, with a key goal of helping physically disadvantaged people like parapalegics have greater control of their bodies.[10] While not exclusively an AI tool, these implants rely critically on AI to identify what patterns correspond to intended movements or commands. Future work should explore the returns to similar AI tools across other disabilities and labor market contexts.

[10]See "Neuralink's brain chip: How brain-computer interfaces may revolutionise treatment for disabilities" in EuroNews.com.

References

Acemoglu, Daron and Joshua D Angrist, "Consequences of Employment Protection? The Case of the Americans with Disabilities Act," *Journal of Political Economy*, 2001, *109* (5), 915–957.

Ameri, Mason, Lisa Schur, Meera Adya, F. Scott Bentley, Patrick McKay, and Douglas Kruse, "The Disability Employment Puzzle: A Field Experiment on Employer Hiring Behavior," *ILR Review*, 2018, *71* (2), 329–364.

Autor, David, Andreas Kostøl, Magne Mogstad, and Bradley Setzler, "Disability benefits, consumption insurance, and household labor supply," *American Economic Review*, 2019, *109* (7), 2613–2654.

Autor, David H. and Mark G. Duggan, "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics*, 02 2003, *118* (1), 157–206.

Baldwin, Marjorie L and William G Johnson, "A critical review of studies of discrimination against workers with disabilities," *Handbook on the Economics of Discrimination*, 2006.

Banerjee, Abhijit, Rema Hanna, Benjamin A Olken, and Diana Sverdlin Lisker, "Social Protection in the Developing World," Working Paper 32382, National Bureau of Economic Research April 2024.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw, "How Does Information Technology Affect Productivity? Plant-level Comparisons of Product Innovation, Process Improvement, and Worker Skills," *Quarterly Journal of Economics*, 2007, *122* (4), 1721–1758.

Bellemare, Charles, Marion Goussé, Guy Lacroix, and Steeve Marchand, "Physical disability and labor market discrimination: evidence from a video résumé field experiment," *American Economic Journal: Applied Economics*, 2023, *15* (4), 452–476.

Benson, Alan, Simon Board, and Moritz Meyer-ter-Vehn, "Discrimination in Hiring: Evidence from Retail Sales," *Review of Economic Studies*, 2024, *Forthcoming.*

Blader, Steven, Claudine Gartenberg, and Andrea Prat, "The Contingent Effect of Management Practices," *Review of Economic Studies*, 2020, *87* (2), 721–749.

Bloom, Nicholas, Gordon B. Dahl, and Dan-Olof Rooth, "Work from Home and Disability Employment," *AER Insights*, 2025.

Bound, John and Richard V. Burkhauser, "Economic Analysis of Transfer Programs Targeted on People with Disabilities," *Handbook of Labor Economics*, 1999, *3*, 3417–3528.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-level evidence," *Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at Work," *Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

Chernozhukov, Victor, Iván Fernández-Val, and Ye Luo, "The Sorted Effects Method: Discovering Heterogeneous Effects Beyond Their Averages," *Econometrica*, 2018, *86* (6), 1911–1938.

Cui, Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers," aug 2025. SSRN working paper, last revised August 21, 2025.

Cullen, Zoe, Will Dobbie, and Mitchell Hoffman, "Increasing the Demand for Workers with a Criminal Record," *Quarterly Journal of Economics*, 2023, *138* (1), 103–150.

DeLeire, Thomas, "The Wage and Employment Effects of the Americans with Disabilities Act," *Journal of Human Resources*, 2000, *35* (4), 693–715.

Deshpande, Manasi, "Does Welfare Inhibit Success? The Long-term Effects of Removing Low-income Youth from the Disability Rolls," *American Economic Review*, 2016, *106* (11), 3300–3330.

_ **and Lee M Lockwood**, "Beyond Health: Nonhealth Risk and the Value of Disability Insurance," *Econometrica*, 2022, *90* (4), 1781–1810.

Forth, John and Melanie K. Jones, "The Disability Pay Gap Within and Across Firms," IZA Discussion Paper 17679, IZA — Institute of Labor Economics, Bonn, Germany February 2025.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, Tobias Kretschmer, and Nick Zubanov, "Is This Really Kneaded? Identifying and Eliminating Potentially Harmful Forms of Workplace Control," *ECONtribute Discussion Paper*, 2025.

Hoffman, Mitchell and Christopher Stanton, "People, Practices, and Productivity: A Review of New Advances in Personnel Economics," *Handbook of Labor Economics*, 2025.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu, "Generative AI and the Nature of Work," Working Paper 25-021 / CESifo WP No. 11479, Harvard Business School; CESifo Working Paper Series April 2024. Distributed for comment and discussion; natural-experiment / regression-discontinuity design using GitHub Copilot deployment.

Krueger, Alan B. and Douglas Kruse, "Labor Market Effects of Spinal Cord Injuries in the Dawn of the Computer Age," NBER Working Paper 5302, National Bureau of Economic Research 1995.

Kruse, Douglas, Alan Krueger, and Lisa Drastal, "Computer Use, Training, and Employment of People with Spinal Cord Injury," *Spine*, 1996, *21* (7), 891–896.

Lalive, Rafael, Jean-Philippe Wuellrich, and Josef Zweimüller, "Do Financial Incentives Affect Firms' Demand for Disabled Workers?," *Journal of the European Economic Association*, 2013, *11* (1), 25–58.

L'Horty, Yannick, Naomie Mahmoudi, Pascale Petit, and François-Charles Wolff, "Is disability more discriminatory in hiring than ethnicity, address or gender? Evidence from a multi-criteria correspondence experiment," *Social Science Medicine*, 2022, *303*, 114990.

Lise, Jeremy, Elena Pastorino, and Luigi Pistaferri, "Revisiting the Employment Effects of the Americans with Disabilities Act," 2023.

Madans, Jennifer H., Julie D. Weeks, and Nazik Elgaddal, "Hearing Difficulties Among Adults: United States, 2019," NCHS Data Brief 414, Centers for Disease Control and Prevention, National Center for Health Statistics July 2021. CDC Stacks archival repository.

Milligan, Kevin, "The Long-Run Growth of Disability Insurance in the United States," in "Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms," University of Chicago Press, 09 2012.

Minor, Dylan, Nicola Persico, and Deborah M. Weiss, "Criminal Background and Job Performance," *IZA Journal of Labor Policy*, 2018, *7*, 1–49.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Oyer, Paul and Scott Schaefer, "Personnel Economics: Hiring and Incentives," *Handbook of Labor Economics*, 2011.

Pallais, Amanda and Emily Glassberg Sands, "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 2016, *124* (6), 1793–1828.

Szerman, Christiane, "The Labor Market Effects of Disability Hiring Quotas," *Available at SSRN 4267622*, 2022.

Wang, Yu, Yang Xie, Minghao Wang, Mengdan Zhao, Rui Gong, Ying Xin, Jia Ke, Ke Zhang, Shaoxing Zhang, Chen Du et al., "Hearing loss prevalence and burden of disease in China: Findings from provincial-level analysis," *Chinese Medical Journal*, 2025, *138* (1), 41–48.

Table 1

Summary Statistics and Tests of Differences Between Disabled and Non-Disabled Workers

	Disabled Workers	Non-Disabled Workers	Full Sample	P-Value of Disabled vs. Non-Disabled Difference
	Mean	Mean	Mean	
	(1)	(2)	(3)	(4)
A. Demographics				
Age	32.82	30.29	30.54	0.00***
% Male Workers	92.55	90.82	90.99	0.00***
B. Registration Type				
% Team Worker	35.36	36.40	36.30	0.18
% Gig Worker	64.64	63.60	63.70	0.18
C. Disability Severity				
% Level 1 (Most Severe)	53.57	-	-	-
% Level 2	16.48	-	-	-
% Level 3	14.24	-	-	-
% Level 4 (Least Severe)	15.71	-	-	-
	N	N	N	
	4,284	39,321	43,605	

Notes: The table is based on a sample of 43,605 food-delivery workers observed between early 2021 and the end of 2023. This includes 4,284 hearing-impaired ("disabled") workers and about 40,000 non-disabled workers. Column (1) shows disabled worker averages; (2) shows non-disabled worker averages; (3) is the full sample average. The unit of observation is worker-level for demographics. Age is the worker's age in years, and % Male Workers is the percentage of workers who are male. % Gig Worker and % Team Worker indicate the worker's registration type. Panel C reports the distribution of disability severity among the disabled workers (Column 1). Severity follows the official four-level classification (GB/T 2634-2010), where Level 1 is the most severe (>90 dB HL) and Level 4 the least severe (41-60 dB HL). The severity information is taken directly from the medical disability certificate submitted by drivers. Column (4) reports the p-value for the difference in means between disabled and non-disabled groups. Asterisks indicate significance levels (*** p<0.01, ** p<0.05, * p<0.1). Disabled workers in this sample tend to be slightly younger (mean age 30.3 vs. 32.8) and less likely to be male (90.8% vs. 92.6%) than non-disabled workers.

Table 2

Descriptive Statistics for Key Outcome Variables (Worker-Week Level)

A. Total Output	Mean	Median	Std. Dev.	**B. Efficiency**	Mean	Median	Std. Dev.	**C. Labor Supply on the Platform**	Mean	Median	Std. Dev.
Orders	128.75	89	141.30	Order → Worker	3.35	2.41	12.35	Online Duration	1519.69	592.51	1807.71
Peak Hour Orders	76.82	52	87.69	Worker → Store	11.59	11.35	7.07	Work Days	4.49	6.00	2.74
Orders (Team)	198.46	194	150.75	Store → Delivery	13.82	13.34	5.59	Bad Weather Work Days	0.63	0.00	1.10
Orders (Gig)	55.01	21	80.40	Late Delivery Rate	81.29	51.15	96.75	Number of Weeks Worked	11.00	3.00	20.00
Restaurants	55.79	47	51.28	Bad Rating	9.62	0.00	94.14				
Customers	116.15	75	127.22								

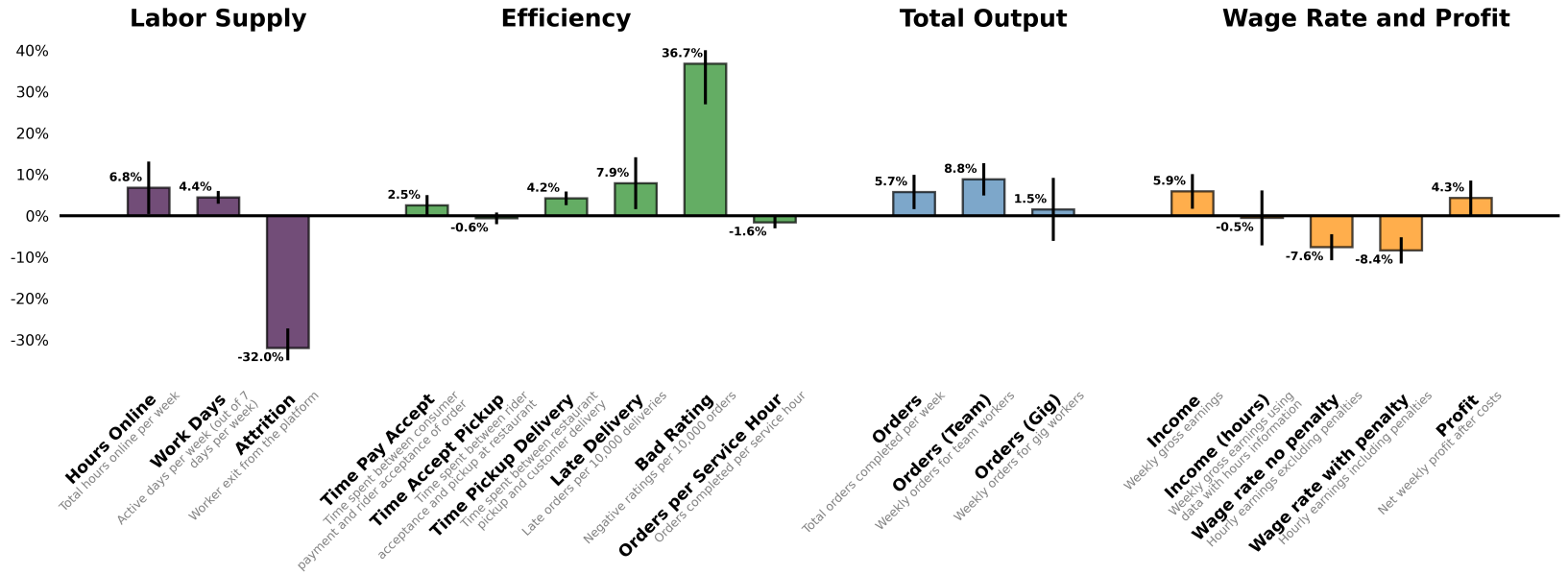
Notes: This table summarizes weekly work outcomes for the delivery workers (with the unit of observation being a worker-week). These statistics are computed from a panel of 568,218 worker-week observations, which comes from tracking 43,605 unique workers over up to 157 weeks (early 2021 through end of 2023). Because many workers did not work every single week (and many left after a short time), the total observed weeks (568k) is less than the theoretical maximum (which would be $43{,}605 \times 157 \approx 6.8$ million).

Panel A: Total Output includes: *Orders* (the total number of delivery orders completed by a worker in a week), *Peak Hour Orders* (the subset of those orders delivered during peak demand hours), *Orders (Team)* and *Orders (Gig)* (weekly orders completed by workers in each registration category – effectively the average weekly orders if the worker is a team worker vs. if they are a gig worker), and the number of unique *Restaurants* and *Customers* served per week.

Panel B: Efficiency records average time durations for each delivery stage: *Order → Worker* is the dispatch time from order placement to a worker accepting it, *Worker → Store* is the travel time from accepting the order to arriving at the restaurant, *Store → Delivery* is the travel time from pickup to delivering to the customer. These times are measured in minutes per order (averaged over the week for each worker). *Late Delivery Rate* is the number of on-time orders, as measured by the pre-set platform standard, normalized per 10,000 orders (to standardize the measure). *Bad Rating* is the number of negative customer ratings received, normalized per 10,000 orders (to standardize the measure).

Panel C: Labor Supply on the Platform covers weekly work effort: *Online Duration* (total time the worker was online and available for orders, measured in minutes per week; 1,520 minutes ≈ 25.3 hours), *Work Days* (number of days the worker was active in that week, out of 7), *Bad Weather Work Days* (days the worker worked in bad weather conditions like rain or snow that week), and *Number of Weeks Worked* which is the total number of weeks the worker has been active on the platform up to that point or in total. (The last variable essentially reflects each worker's tenure or longevity on the platform, with median 3 weeks and mean 11 weeks, indicating high turnover.)

Figure 1: Pre-AI Normalized Differences Between Disabled and Non-Disabled Workers in Labor Supply, Efficiency, and Output



Notes: This Figure visually summarizes the performance gaps between disabled and non-disabled workers across key metrics before the intervention, based on regression results from Table A.2. Each bar represents a percentage difference associated with being a disabled worker for a given outcome (e.g., weekly orders, delivery time).

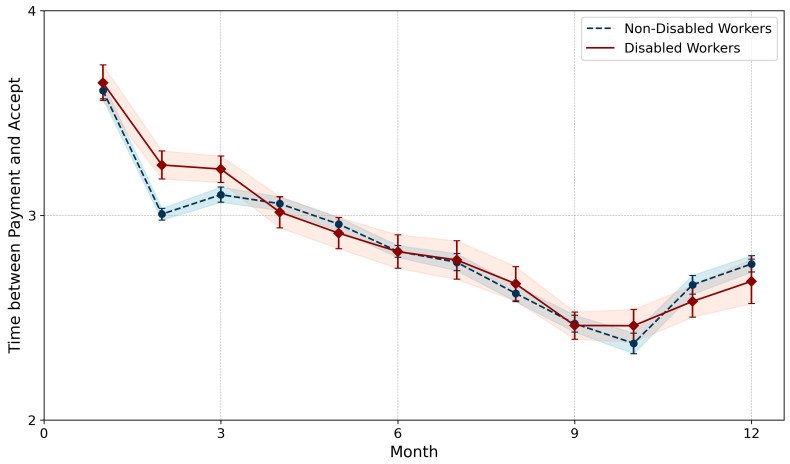
Labor Supply covers weekly work effort: *Online Duration* (total time the worker was online and available for orders, measured in minutes per week), *Work Days* (number of days the worker was active in that week, out of 7), *Bad Weather Work Days* (days the worker worked in bad weather conditions like rain or snow that week), and *Survival* reflects each worker's tenure or longevity on the platform.

Efficiency records average time durations for each delivery stage: *Time Pay Accept* is the dispatch time from order placement to a worker accepting it, *Time Accept Pickup* is the travel time from accepting the order to arriving at the restaurant, *Time Pickup Delivery* is the travel time from pickup to delivering to the customer. These times are measured in minutes per order (averaged over the week for each worker). *Bad Rating* and *Late Delivery* represent negative customer ratings and orders delivered after the scheduled arrival time, respectively, each normalized per 10,000 orders.

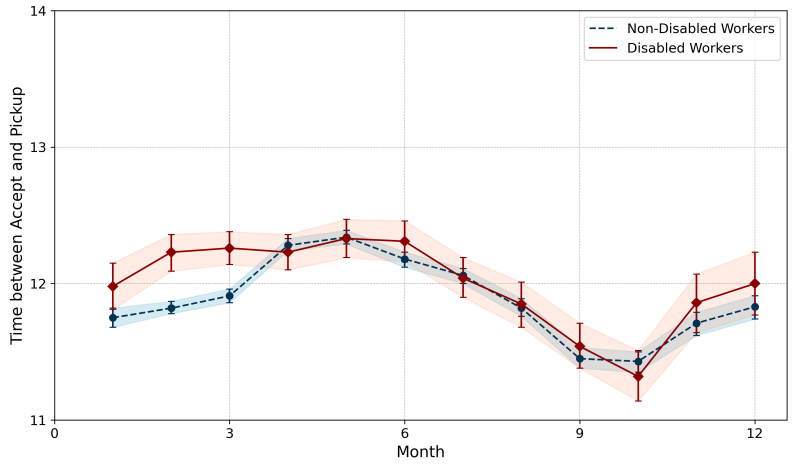
Total Output includes: *Orders* (the total number of delivery orders completed by a worker in a week), *Peak Hour Orders* (the subset of those orders delivered during peak demand hours), *Orders (Team)* and *Orders (Gig)* (weekly orders completed by workers in each registration category – effectively the average weekly orders if the worker is a team worker vs. if they are a gig worker), and the number of unique *Restaurants* and *Customers* served per week.

Wage Outcomes includes *Income* (weekly gross earnings), *Wage* (average earnings per hour online), and *Profit* (net weekly profit for the platform after deducting measured operating costs).

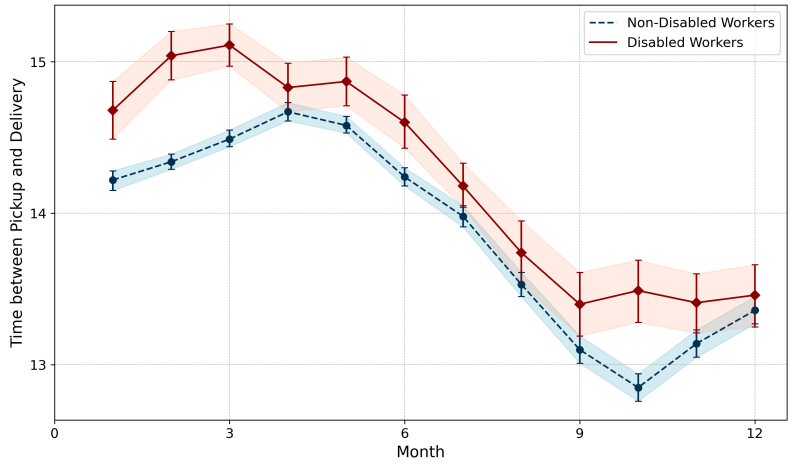
Figure 2: Pre-AI Tenure-Productivity Curves: Time Spent on Different Parts of the Order on Months of Tenure



(a) Time Spent Between Order → Worker (Per Order)



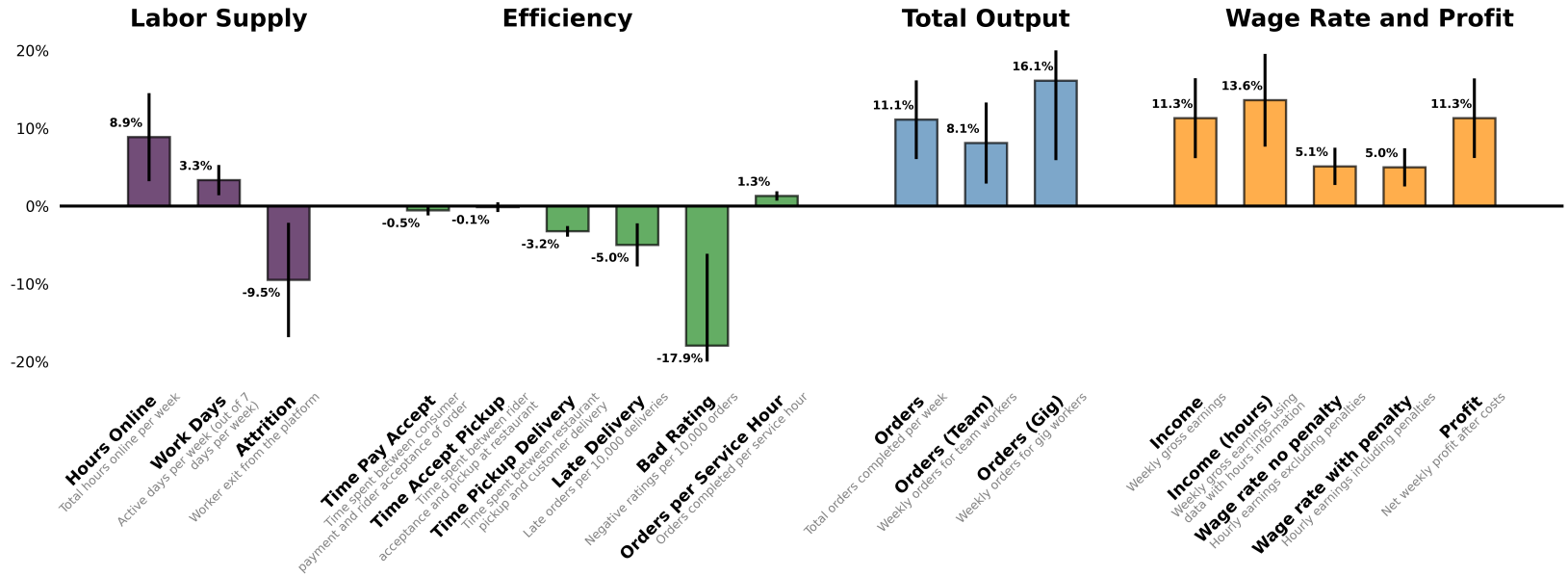
(b) Time Spent Between Worker → Store (Per Order)



(c) Time Spent Between Store → Delivery (Per Order)

Notes: Figure 2 examines how worker efficiency (speed) changes with experience (*tenure*) on the platform. It plots the average time spent per order at each month of a new worker's tenure for both disabled and non-disabled groups across three panels: **Panel (a) Order → Worker**: Average dispatch time (minutes per order) from customer order to rider assignment, tracked by tenure month. **Panel (b) Worker → Store**: Average time (minutes per order) from assignment to arrival at the restaurant. **Panel (c) Store → Delivery**: Average time (minutes per order) from food pickup to customer delivery. Each point on a curve represents the mean time for that stage among workers with a given tenure month. For example, at "Month 1," the figure shows the average per-order time for all workers in their first month, while at "Month 6," it represents the average for those with six months of experience. Error bars represent the 5th to 95th percentile confidence interval. The red solid line is for disabled workers and the blue dashed line is for non-disabled workers.

Figure 3: The Impact of the AI Tool on the Outcomes of Disabled Workers



Notes: Figure 3 is analogous to Figure 1 but focuses on the impact of the Intelligent Outbound Call intervention. This visualization is based on the difference-in-differences regression results (Table A.3). Each bar represents the shift in the disabled vs. non-disabled gap due to the intervention, expressed in percentage change. Positive bars indicate an improvement from the disabled workers' perspective, such as an increase in orders (beneficial) or potentially longer working hours. Negative bars suggest a decline in the metric for disabled workers relative to non-disabled workers. However, for time-related or bad review metrics, a negative bar actually signifies improvement, as it reflects a reduction in time spent or complaints received. Error bars represent confidence intervals; if they cross zero, the effect may not be statistically significant.

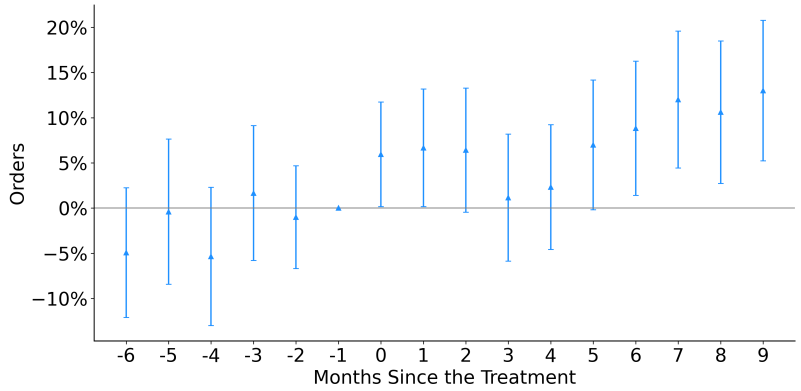
Labor Supply covers weekly work effort: *Online Duration* (total time the worker was online and available for orders, measured in minutes per week), *Work Days* (number of days the worker was active in that week, out of 7), *Bad Weather Work Days* (days the worker worked in bad weather conditions like rain or snow that week), and *Survival* reflects each worker's tenure or longevity on the platform.

Efficiency records average time durations for each delivery stage: *Time Pay Accept* is the dispatch time from order placement to a worker accepting it, *Time Accept Pickup* is the travel time from accepting the order to arriving at the restaurant, *Time Pickup Delivery* is the travel time from pickup to delivering to the customer. These times are measured in minutes per order (averaged over the week for each worker). *Bad Rating* and *Late Delivery* represent negative customer ratings and orders delivered after the scheduled arrival time, respectively, each normalized per 10,000 orders.

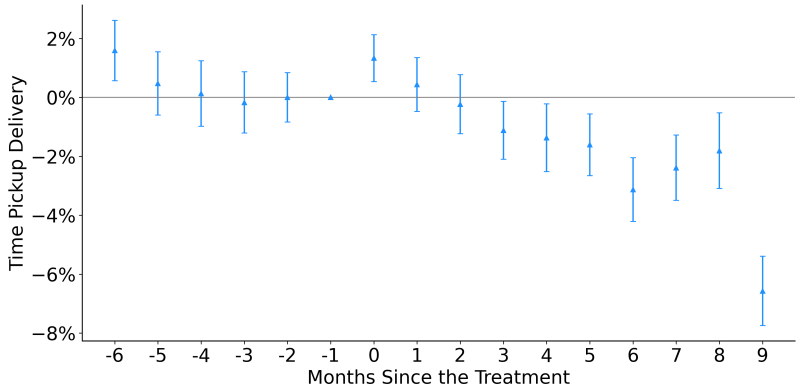
Total Output includes: *Orders* (the total number of delivery orders completed by a worker in a week), *Peak Hour Orders* (the subset of those orders delivered during peak demand hours), *Orders (Team)* and *Orders (Gig)* (weekly orders completed by workers in each registration category – effectively the average weekly orders if the worker is a team worker vs. if they are a gig worker), and the number of unique *Restaurants* and *Customers* served per week.

Wage Outcomes includes *Income* (weekly gross earnings), *Wage* (average earnings per hour online), and *Profit* (net weekly profit for the platform after deducting measured operating costs).

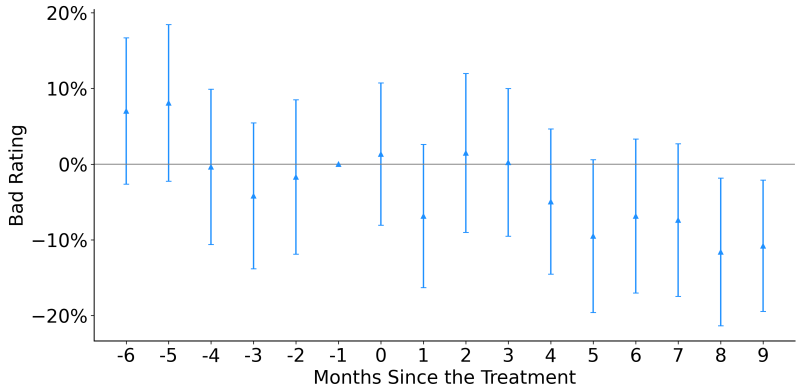
Figure 4: Event Study Analysis: Productivity



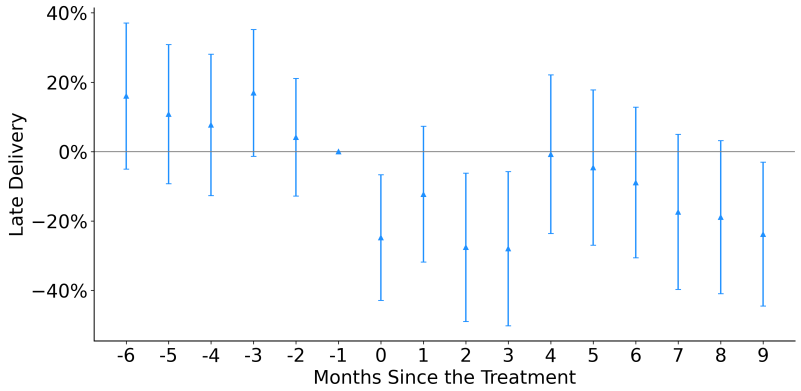
(a) Number of Orders Completed



(b) Time Spent Between Store → Delivery (Per Order)



(c) Bad Ratings



(d) Number of Delayed Orders

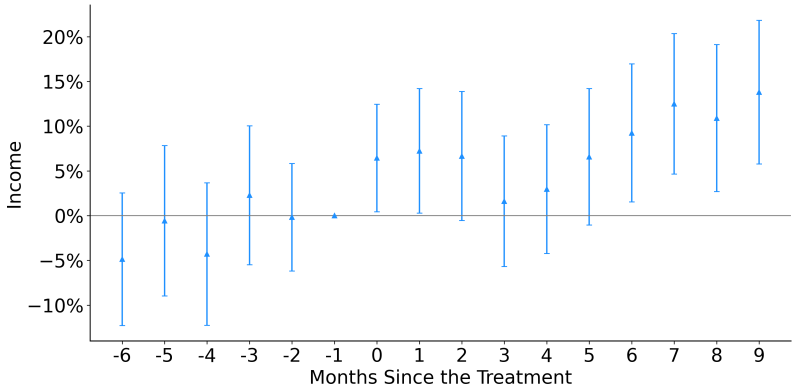
Notes: Figure 4 presents an event study analysis of the introduction of the Intelligent Outbound Call (IOC) system, examining trends before and after the intervention for outcome measures related to productivity.

Panel (a) plots the difference in weekly orders between disabled and non-disabled workers for each time period relative to the IOC implementation. The x-axis represents the number of months relative to the intervention, with 0 indicating the implementation week, while the y-axis shows the estimated difference in orders between disabled and non-disabled workers.

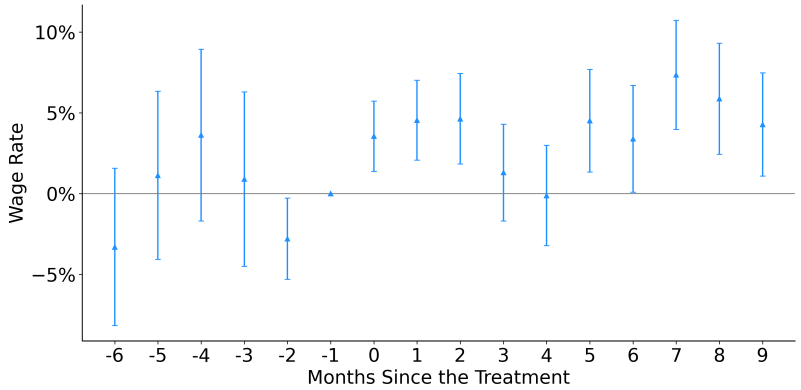
Panel (b) shows the gap in delivery time for the final stage (*Store → Customer*) between disabled and non-disabled workers, plotted relative to the implementation of IOC.

Panel (c) and (d) report differences in the incidence of *Negative Customer Ratings* and *Late Deliveries*, each normalized per 10,000 orders.

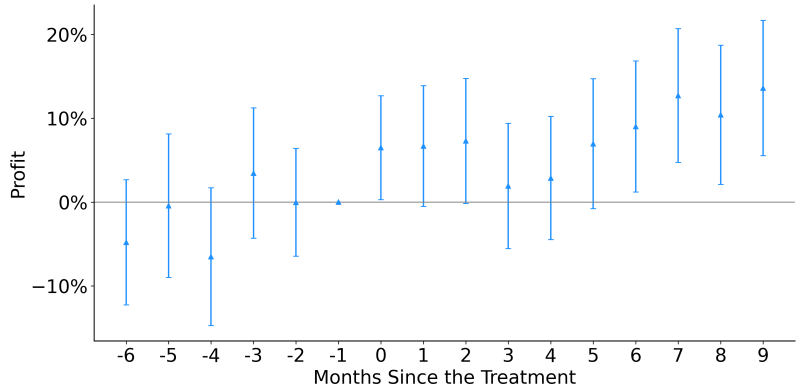
Figure 5: Event Study Analysis: Wage and Profits



(a) Income



(b) Wage



(c) Profit

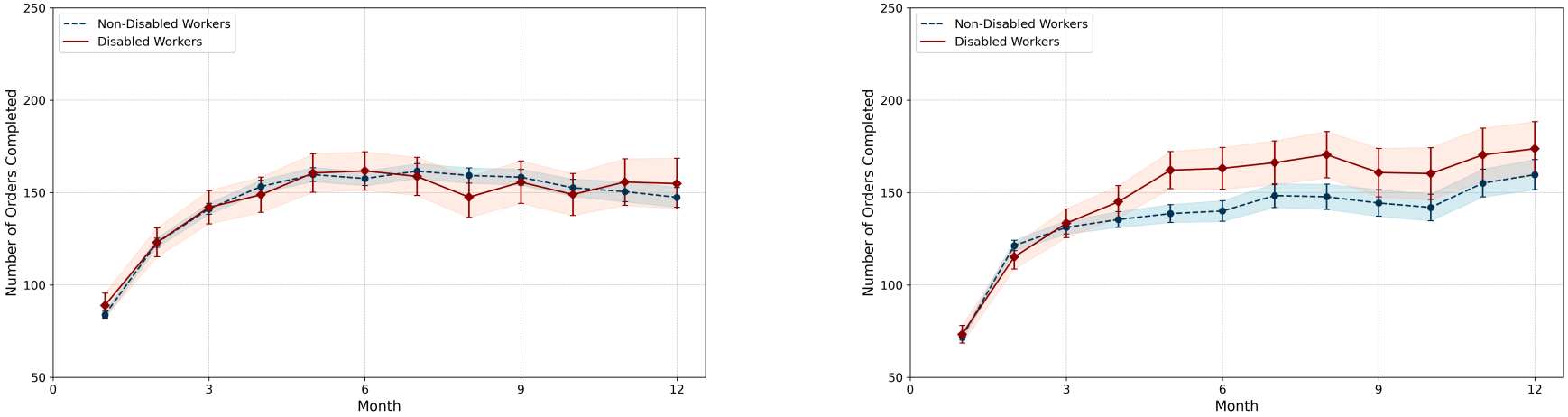
Notes: Figure 5 presents an event study analysis of the introduction of the Intelligent Outbound Call (IOC) system, examining trends before and after the intervention for outcome measures related to wages and profits.

Panel (a) plots difference in income (measured by weekly gross earnings) between disabled and non-disabled workers across time periods relative to IOC implementation. The x-axis indicates months relative to the intervention, with 0 marking the implementation week, and the y-axis shows the estimated difference in gross income.

Panel (b) shows Difference in wage (measured by average earnings per hour online).

Panel (c) shows Difference in profit generated (measured by net weekly profit for the platform after deducting operating costs).

Figure 6: Tenure-Productivity Curves on Number of Orders Completed, Before and After AI



(a) Pre-Periods

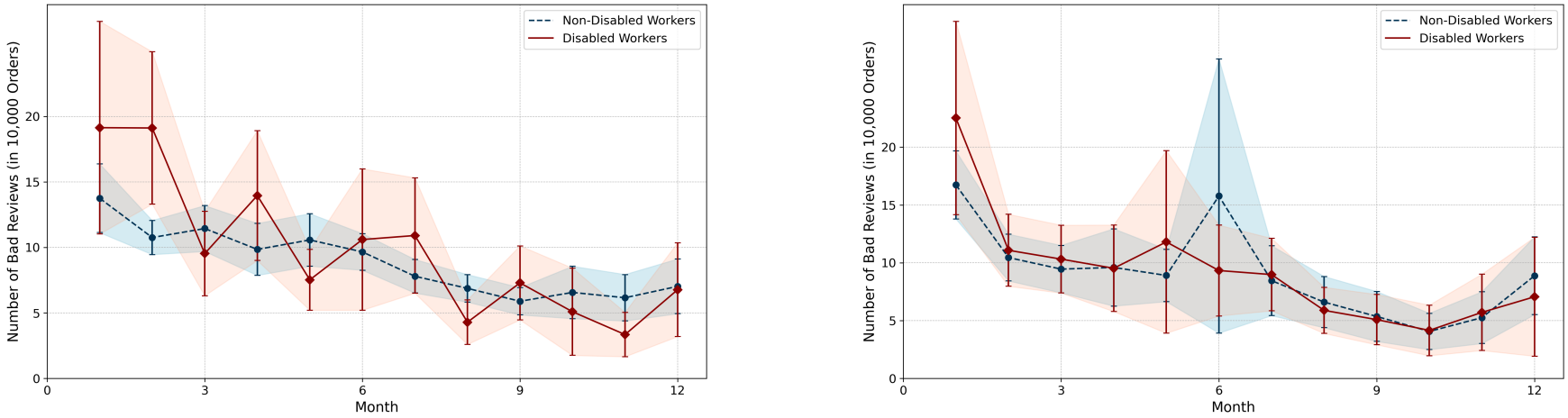
(b) Post-Periods

Notes: Figure 6 examines how the productivity trajectory of new workers over their first year differs before and after the introduction of the Intelligent Outbound Call (IOC) technology. It focuses on the number of orders completed per month of tenure. Error bars represent the 5th to 95th percentile confidence interval. The red solid line represents disabled workers, while the blue dashed line represents non-disabled workers. The tenure curves follow a similar approach to Figure 2 (cohort analysis or regression by tenure month) but are split based on whether workers had access to IOC from the start. The IOC system was implemented in April 2022, with the pre-period covering January 2021 to early 2022, and the post-period starting in April 2022 and extending 12 months onward.

Panel (a) Pre-Period: Displays the tenure curve of monthly order counts for disabled and non-disabled workers who started before IOC was implemented (i.e., cohorts from the pre-IOC era). For each tenure month (1–12), it plots the average number of orders completed by workers of that tenure.

Panel (b) Post-Period: Shows the corresponding tenure curves for workers who started after IOC was introduced (the post-IOC era). It tracks the monthly deliveries of disabled and non-disabled workers from month 1 to 12.

Figure 7: Tenure Curves: Bad Reviews (in 10,000 Orders)



(a) Pre-Periods

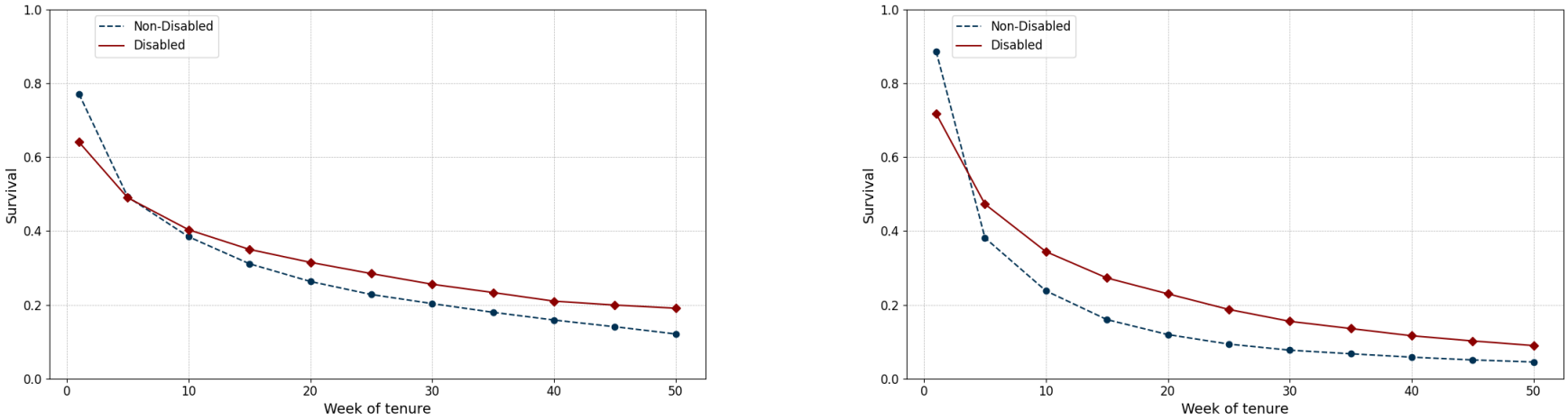
(b) Post-Periods

Notes: Figure 7 is analogous to Figure 6, but focuses on service quality, measured by negative customer reviews. It illustrates how the rate of bad reviews changes as workers gain experience and how this trajectory differed before and after the introduction of IOC. The metric used is the number of bad reviews per 10,000 orders for each month of tenure. Error bars represent the 5th to 95th percentile confidence interval. The red solid line represents disabled workers, while the blue dashed line represents non-disabled workers. The IOC system was implemented in April 2022, with the pre-period covering January 2021 to early 2022, and the post-period beginning in April 2022 and extending 12 months onward.

Panel (a) Pre-Period: Displays the tenure curve of bad review rates for new disabled and non-disabled workers who joined before IOC was introduced. For each month of experience (1–12), it plots the average negative review rate per 10,000 orders for those workers.

Panel (b) Post-Period: Shows the corresponding tenure curves for cohorts who started after IOC.

Figure 8: Survival Curves: Before and After Comparison



(a) Pre-Periods

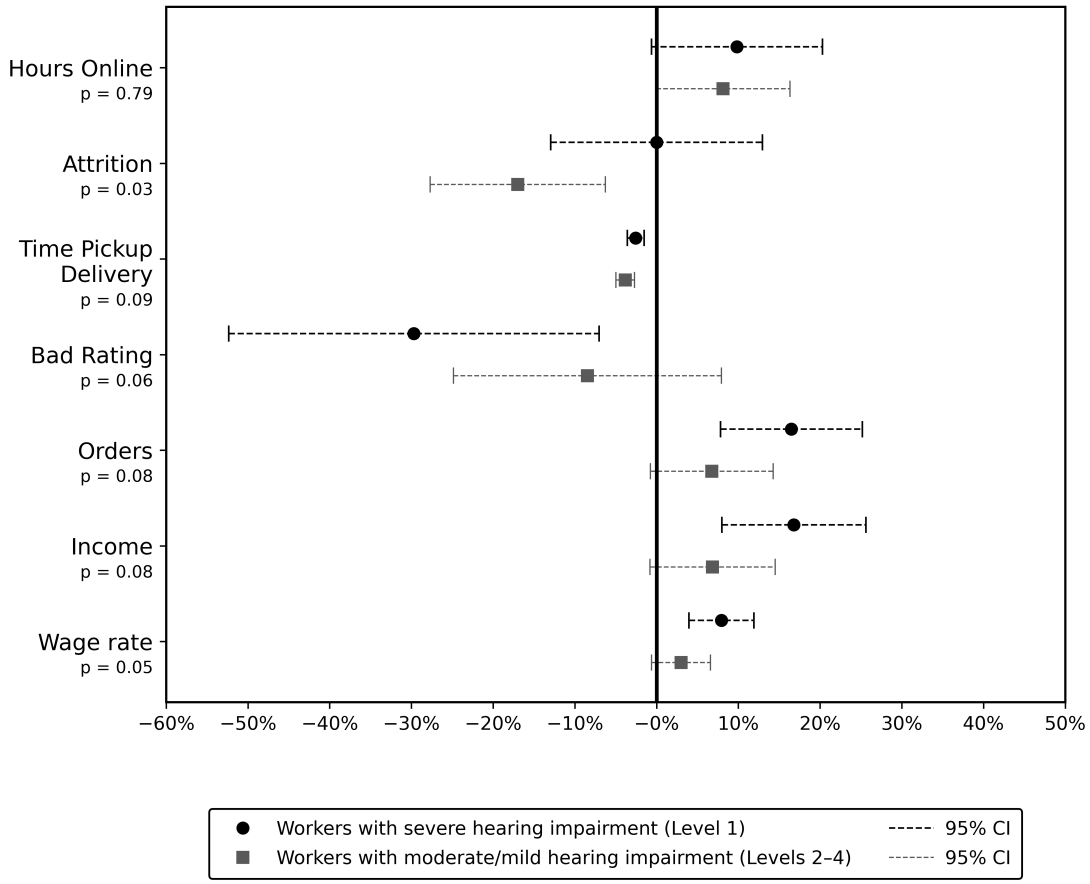
(b) Post-Periods

Notes: Figure 8 compares the retention (survival) of workers on the platform for cohorts before and after the introduction of the Intelligent Outbound Call (IOC) system, highlighting differences between disabled and non-disabled workers. The y-axis represents the fraction of workers still active (i.e., those who have not quit), while the x-axis shows the number of weeks since a worker started. The red solid line represents disabled workers, while the blue dashed line represents non-disabled workers. The IOC system was implemented in April 2022. The pre-period covers January 2021 to early 2022, while the post-period begins in April 2022 and extends 12 months onward. The starting point at week 0 (or week 1) is 100% by definition, as all workers are initially registered as potential workers. As weeks progress, the curves decline as workers leave the platform.

Panel (a) presents the survival curves for disabled and non-disabled workers who joined before the IOC system was introduced. It illustrates the percentage of workers remaining active over time (week 1, week 2, etc.). The week 1 result shows the number of workers who registered and delivered at least one order in their first week: approximately 22% of non-disabled registrants never actually begin the job (i.e., they sign up but never complete their first order), while for disabled workers, this figure is much higher at 38%. This gap suggests that disabled workers are more likely to drop out at the very beginning, possibly due to concerns about communication barriers or a lack of onboarding support. However, among those who do start working, disabled workers tend to have higher ongoing retention than non-disabled workers. Their survival curve declines more gradually, indicating that they remain active for longer.

Panel (b) displays the corresponding survival curves for workers who started after IOC was implemented. The results show that the initial onboarding drop is still present even after the technology change, since IOC mainly helps once workers start delivering. It might not affect someone's decision to start in the first place, as that decision could depend on other factors like confidence or training. However, except for this initial drop, more disabled workers remain active at each subsequent point compared to Panel (a). The gap between disabled and non-disabled survival widened in the post-IOC cohort, meaning that the technology not only improved performance but also encouraged disabled couriers to remain on the job longer.

Figure 9: Heterogeneous Treatment Effects by Disability Severity



Notes: Figure 9 reports heterogeneous treatment effects by disability severity. Severity follows the official four-level classification (GB/T 2634-2010), where Level 1 is the most severe (>90 dB HL) and Level 4 the least severe (41-60 dB HL). The severity information is taken directly from the medical disability certificate submitted by drivers.

Online Appendices (Not for Publication)

Empowering Inclusive Work

by Yanyou Chen, Mitchell Hoffman, Huilan Xu, and Zhe Yuan

A Supplementary Empirical Results

Table A.1: Pre-AI Gaps in Wages and Profits, and How are these Affected by AI

	(1)	(2)	(3)
	Inverse Hyp. Sine of Weekly Income	Inverse Hyp. Sine of Hourly Pay	Inverse Hyp. Sine of Platform Profits
Panel A: Pre-AI Differences			
Disabled	0.0443**	-0.0846***	0.0335
	(0.0225)	(0.0160)	(0.225)
Observations	181,539	66,672	181,539
Panel B: Impact of AI			
Disabled × Post	0.114***	0.0510***	0.113***
	(0.0217)	(0.0114)	(0.0214)
Observations	468,346	332,482	468,346

Notes: An observation is a worker-week. Standard errors clustered by worker in parentheses. Hourly Pay (Column 2) is calculated as income divided by online duration. There are two main reasons for the smaller number of observations in Column (2). First, online duration records are missing for all workers during the first four months of the sample, and approximately 50% of records are missing between months 5 and 16, immediately prior to the AI intervention. Afterward, the fraction of missing records falls below 1%. Second, we exclude extreme values from the analysis, defined as driver–weeks with more than 10 orders per hour, fewer than 1 order per hour, or a wage rate exceeding 70 RMB (approximately 10 USD). We exclude approximately 5–10% of observations in each period, and this exclusion rate remains consistent over time. The incidence of missing records and excluded extreme values is balanced across disabled and non-disabled workers throughout the sample period. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A.2

Pre-AI Differences Between Disabled and Non-Disabled Workers

A. Total Output	Estimate	Observations	R-squared	**B. Efficiency**	Estimate	Observations	R-squared	**C. Labor Supply**	Estimate	Observations	R-squared
Orders	2.336***	206,966	0.341	Order → Worker	0.155***	181,539	0.197	Online Duration	24.62***	206,966	0.306
	(0.761)				(0.0232)				(7.070)		
Peak Hour Orders	1.909***	206,966	0.332	Worker → Store	0.205***	181,539	0.093	Work Days	0.235***	206,966	0.253
	(0.472)				(0.0319)				(0.0165)		
Orders (Team)	9.786***	116,370	0.108	Store → Delivery	0.820***	181,539	0.326	Bad Weather Work Days	0.0432***	206,966	0.168
	(1.225)				(0.0360)				(0.00753)		
Orders (Gig)	-4.810***	90,596	0.043	Late Delivery Rate	34.24***	181,539	0.034	Survival	0.0628***	15,710	0.108
	(0.816)				(5.867)				(0.0130)		
Restaurants	4.320***	206,966	0.243	Bad Rating	5.183***	181,539	0.006				
	(0.303)				(0.684)						
Customers	6.484***	206,966	0.304								
	(0.730)										

Notes: This table presents regression results comparing the performance of disabled and non-disabled workers before the introduction of the communication tool. Each row represents an outcome variable, with the reported coefficient indicating the estimated difference associated with the disabled worker indicator. The regressions use data from January 2021 to mid-April 2022 (pre-policy period). The regression includes various controls: worker role (team vs. gig), age, and gender, along with fixed effects for registration year, week of observation, and city of operation. We also include the time fixed effect. Standard errors, clustered at the individual level, are reported in parentheses. Asterisks denote statistical significance: (*** $p < 0.01$).

Panel A: Total Output – Weekly Orders completed, Peak Hour Orders, and the breakdown by type (*Orders (Team)* and *Orders (Gig)*), as well as unique *Restaurants* and *Customers* served per week. These are the same definitions as in Table 2, but here we focus on the difference for disabled workers. For example, the coefficient for "Orders" (2.336***) means disabled workers complete about 2.336 more orders per week than similar non-disabled workers on average.

Panel B: Efficiency – Average time per order in minutes for each stage: *Order → Worker* (order dispatch time), *Worker → Store* (travel time to the restaurant), *Store → Delivery* (travel time to the customer), *Late Delivery Rate* (the number of over time orders per 10,000 orders) and *Bad Rating* (the number of bad customer reviews per 10,000 orders). A positive coefficient for a time means disabled workers take longer (which is less efficient), while a positive coefficient for *Late Delivery Rate* and *Bad Rating* means disabled workers receive more negative reviews than non-disabled workers.

Panel C: Labor Supply & Retention – *Online Duration* (total weekly time online, in minutes), *Work Days* (days worked per week), *Bad Weather Work Days* (days worked in weeks with bad weather), and *Survival*. Here, "Survival" is a measure of retention: it indicates the difference in the probability of remaining active on the platform for disabled vs. non-disabled workers. In this context, the survival coefficient (0.0628***) can be interpreted as disabled workers being 0.0628 higher in the survival metric – roughly a 6.3 percentage point higher chance of staying on the platform at any given time compared to non-disabled workers.

Table A.3

Treatment Effects of the Intelligent Outbound Call (Difference-in-Differences Results)

A. Total Output	Estimate	Observations	R-squared	**B. Efficiency**	Estimate	Observations	R-squared	**C. Labor Supply**	Estimate	Observations	R-squared
Orders	5.800***	559,888	0.638	Order → Worker	0.0551	480,070	0.007	Online Duration	-73.22	559,888	0.678
	(1.545)				(0.112)				(101.3)		
Peak Hour Orders	2.500***	559,888	0.634	Worker → Store	0.0331	480,070	0.000	Work Days	0.138*	559,888	0.530
	(0.780)				(0.0576)				(0.0704)		
Orders (Team)	2.735*	289,786	0.502	Store → Delivery	-0.291***	480,070	0.001	Bad Weather Work Days	0.0489***	559,888	0.294
	(1.545)				(0.0565)				(0.0137)		
Orders (Gig)	9.216***	270,102	0.568	Late Delivery Rate	-10.13***	480,070	0.502	Survival	0.062***	43,650	0.109
	(1.221)				(3.584)				(0.00986)		
Restaurants	3.771***	559,888	0.675	Bad Rating	-2.162**	480,070	0.351				
	(0.429)				(1.073)						
Customers	6.108***	559,888	0.663								
	(1.085)										

Notes: This analysis uses the full panel from 2021 to 2023 (156 weeks), covering both pre- and post-intervention periods. The IOC technology was introduced in late April 2022 (week 17 of 2022); "Pre-Periods" refer to weeks before this date, and "Post-Periods" to those after. The regression is at the worker-week level, including both disabled and non-disabled workers. All regression analyses account for the fixed effects of the week and worker. Robust standard errors are clustered at worker level are and reported in parentheses. *** p<0.01. The table reports the impact of the new assistive technology on these outcomes. Each coefficient represents the estimated additional change for disabled workers following the IOC feature's introduction, relative to any changes among non-disabled workers. This difference-in-differences estimate captures how the disabled–non-disabled gap evolved post-intervention.

The structure remains consistent as in Table A.2:

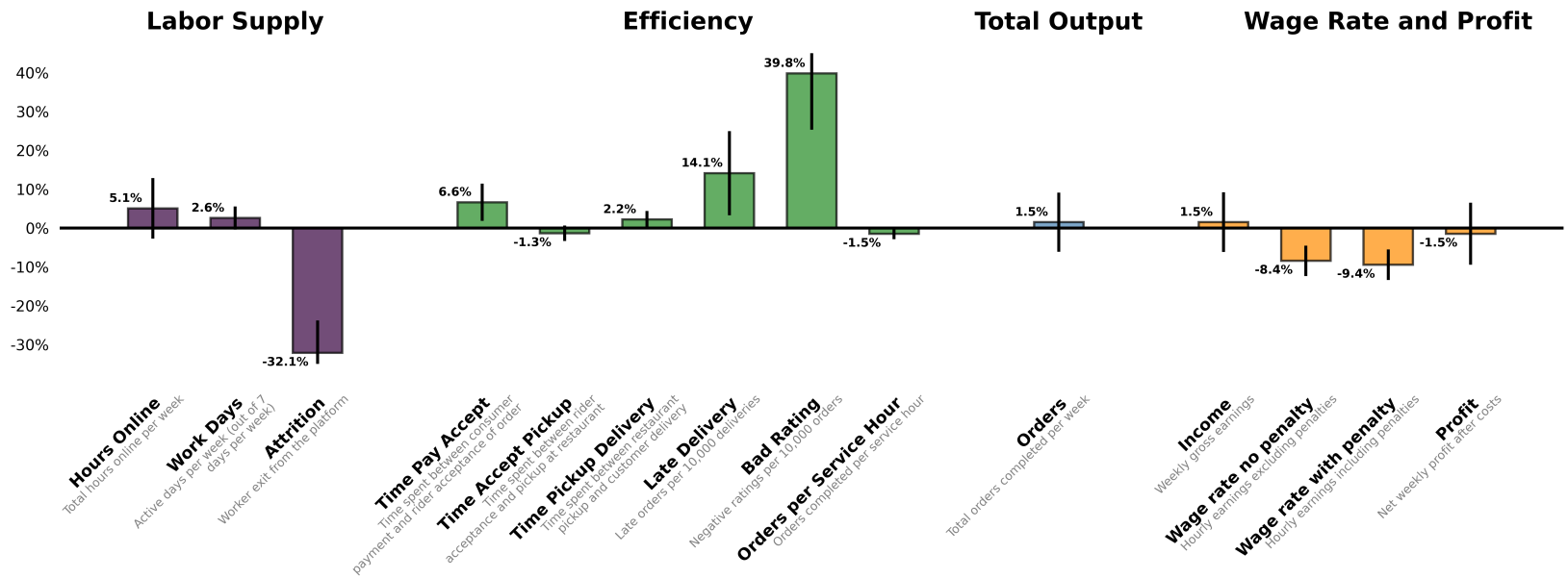
Panel A: Total Output (weekly orders, peak-hour orders, orders by team/gig type, and unique restaurants and customers served).

Panel B: Efficiency (order processing times, over time orders and bad review rate).

Panel C: Labor Supply & Retention (online hours, workdays, bad-weather days, and retention/survival metrics).

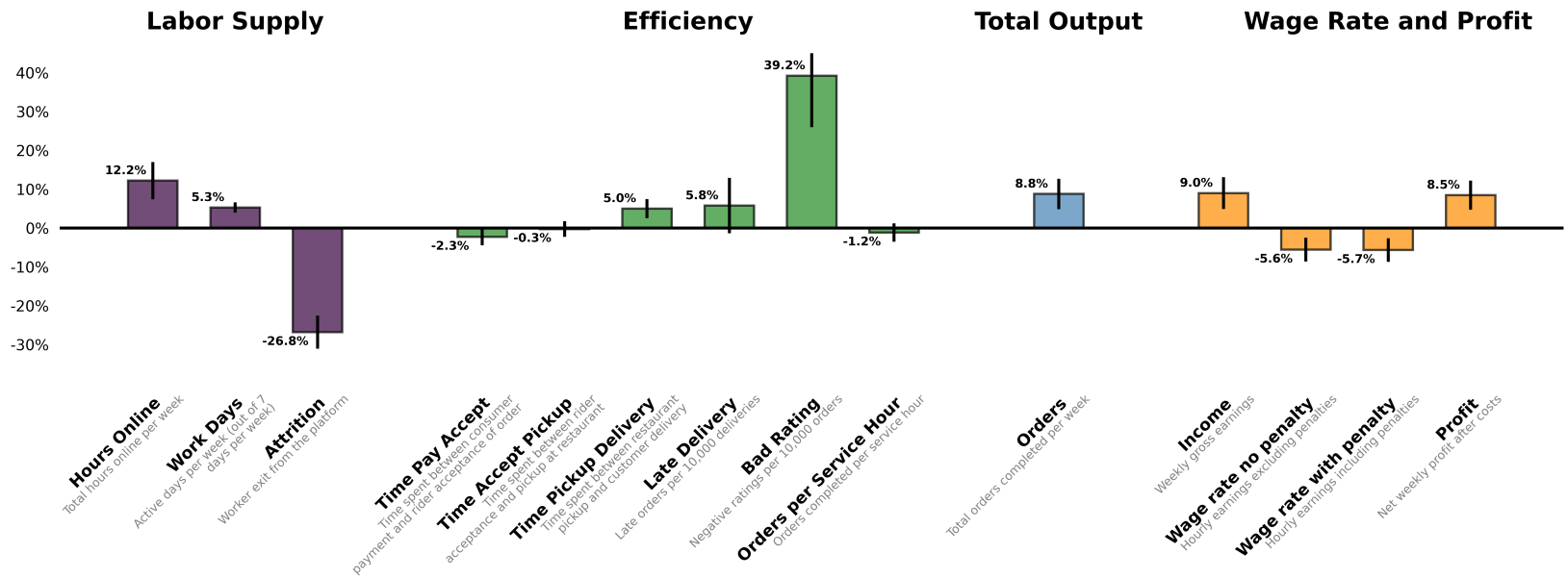
For example, a coefficient of 5.800*** on "Orders" indicates that after the communication tool was introduced, disabled workers completed about 5.8 more orders per week than they would have otherwise, accounting for trends in the control group. A negative coefficient, such as -0.291*** for "Store → Delivery" time, suggests that disabled workers' delivery time per order decreased by 0.291 minutes relative to the control group, indicating an efficiency improvement due to the tool.

Figure A.1: Pre-AI Differences Between Disabled and Non-Disabled Workers: Gig Workers Only



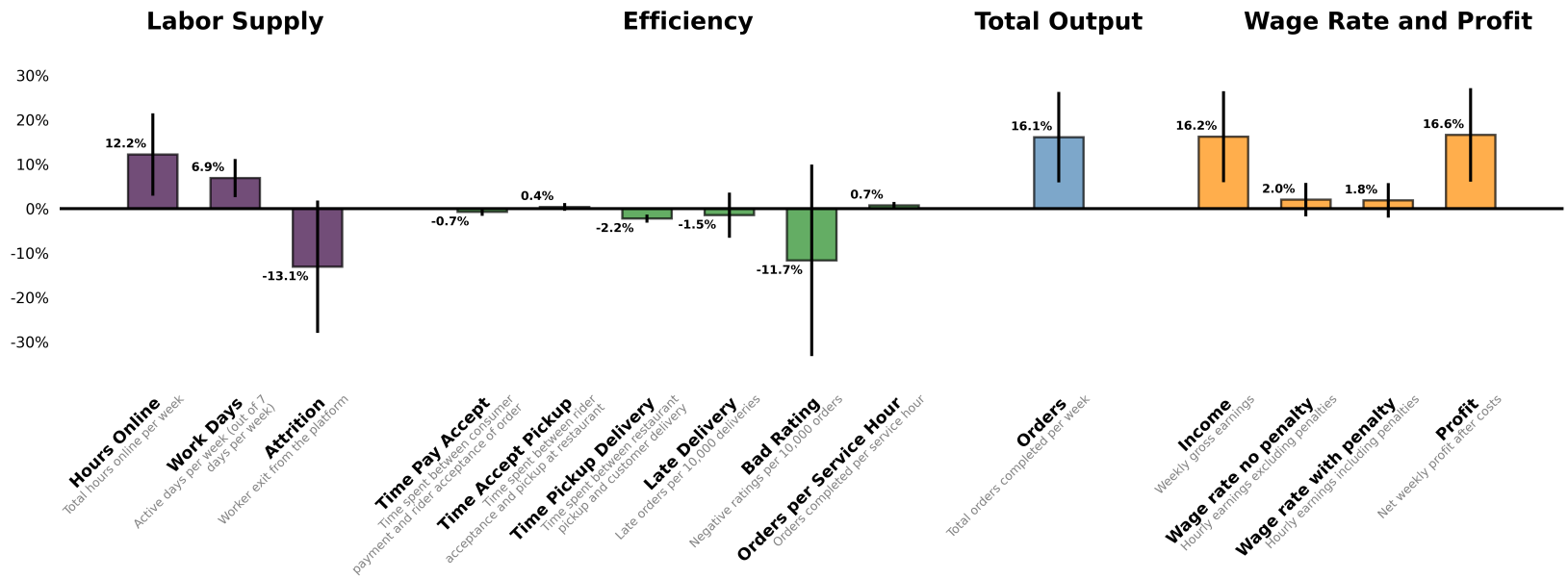
Notes: This figure reproduces the analysis in Figure 1, focusing exclusively on gig workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept.

Figure A.2: Pre-AI Differences Between Disabled and Non-Disabled Workers: Team Workers Only



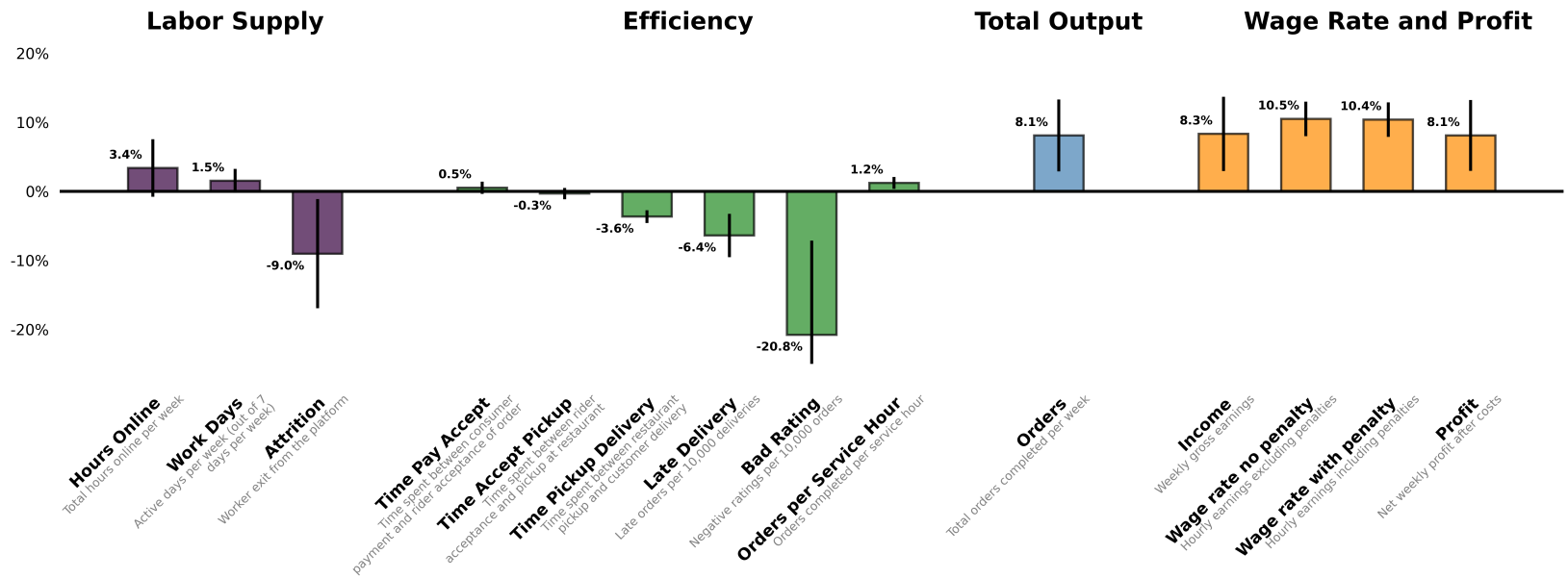
Notes: This figure replicates the analysis in Figure 1, but restricts the sample to team workers only. As described in the main text, team workers are assigned to fixed stations and typically follow a structured wage scheme, where unit prices increase stepwise with the number of monthly completed orders. Their pickup assignments are determined by the platform's algorithm, with priority given to workers who have completed more orders in the past, while disability status is not considered.

Figure A.3: Impact of the AI Tool on Disabled Worker Outcomes: Gig Workers Only



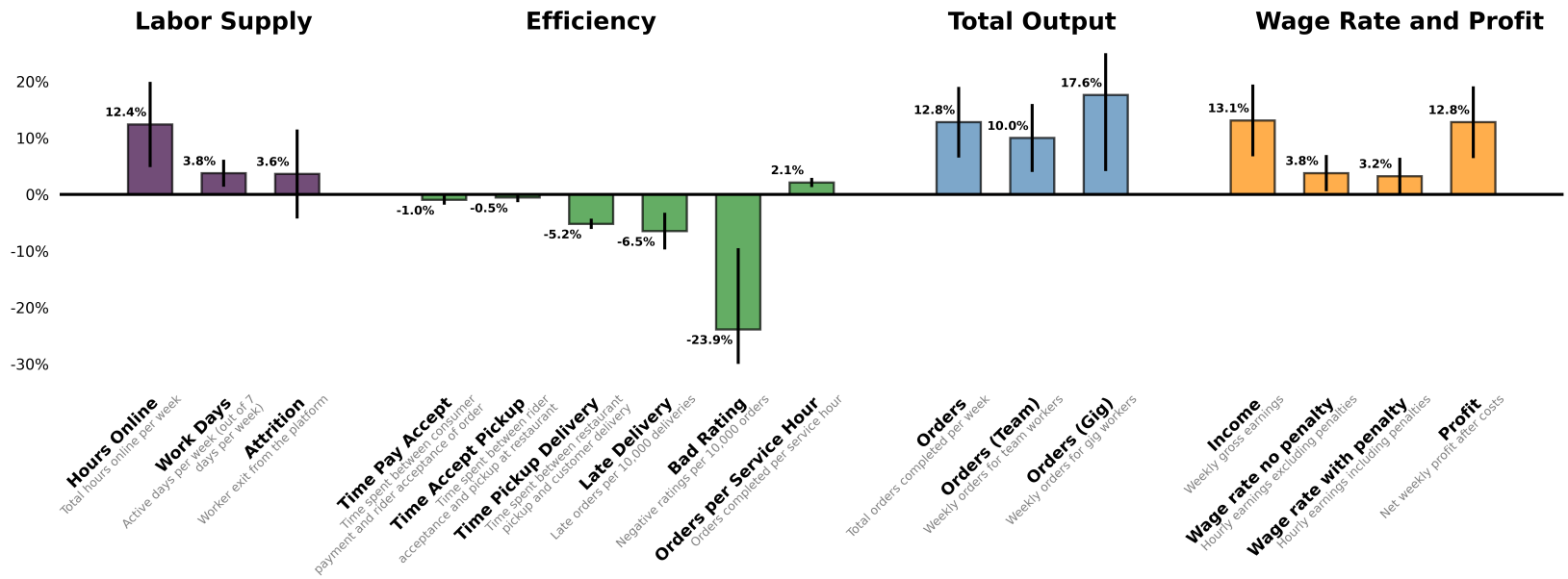
Notes: This figure reproduces the analysis in Figure 3, focusing exclusively on gig workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept.

Figure A.4: Impact of the AI Tool on Disabled Worker Outcomes: Team Workers Only



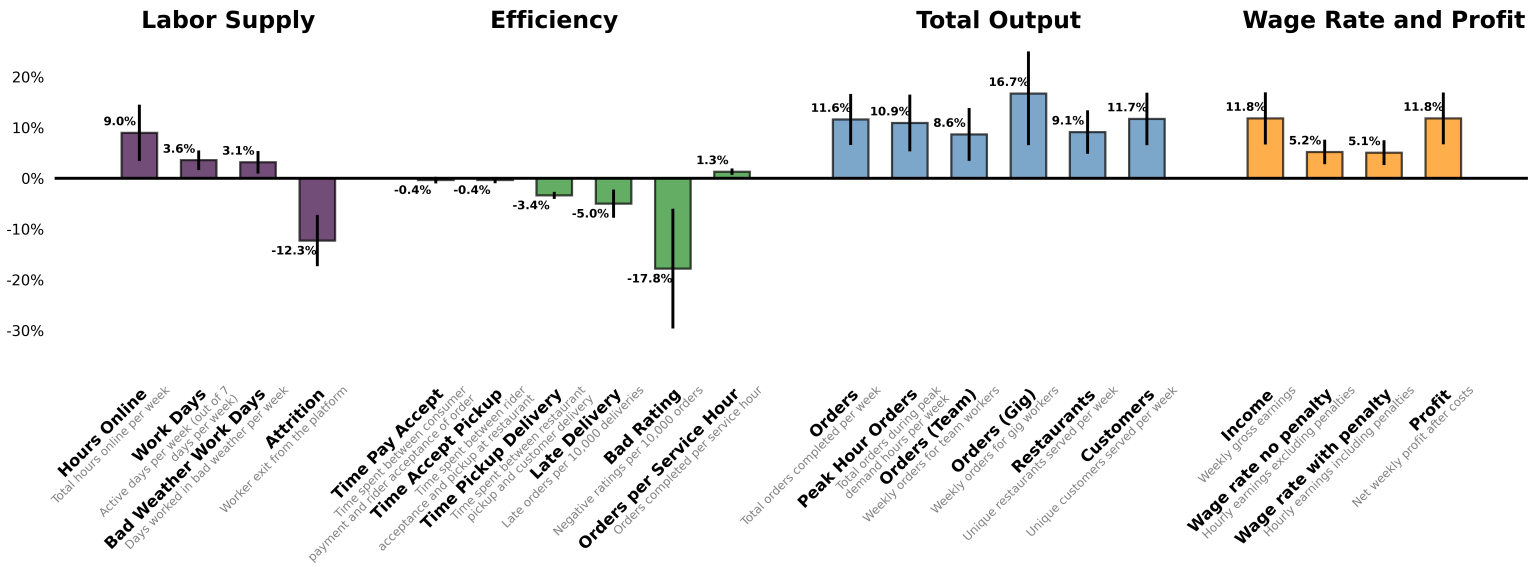
Notes: This figure replicates the analysis in Figure 3, but restricts the sample to team workers only. As described in the main text, team workers are assigned to fixed stations and typically follow a structured wage scheme, where unit prices increase stepwise with the number of monthly completed orders. Their pickup assignments are determined by the platform's algorithm, with priority given to workers who have completed more orders in the past, while disability status is not considered.

Figure A.5: Threat to Identification: Excluding the First Four Months of Orders



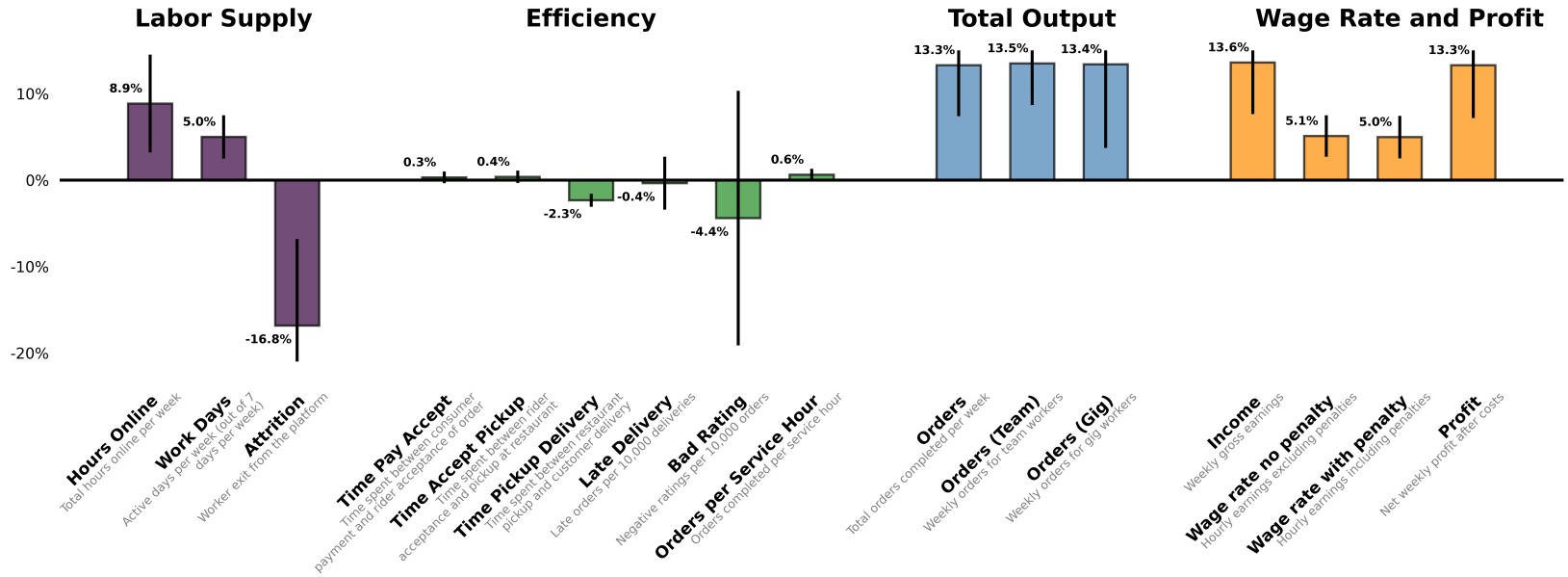
Notes: This figure replicates the analysis in Figure 3, but restricts the sample excluding the first four months of orders.

Figure A.6: Threat to Identification: Restrict to Workers Present Pre- & Post-AI



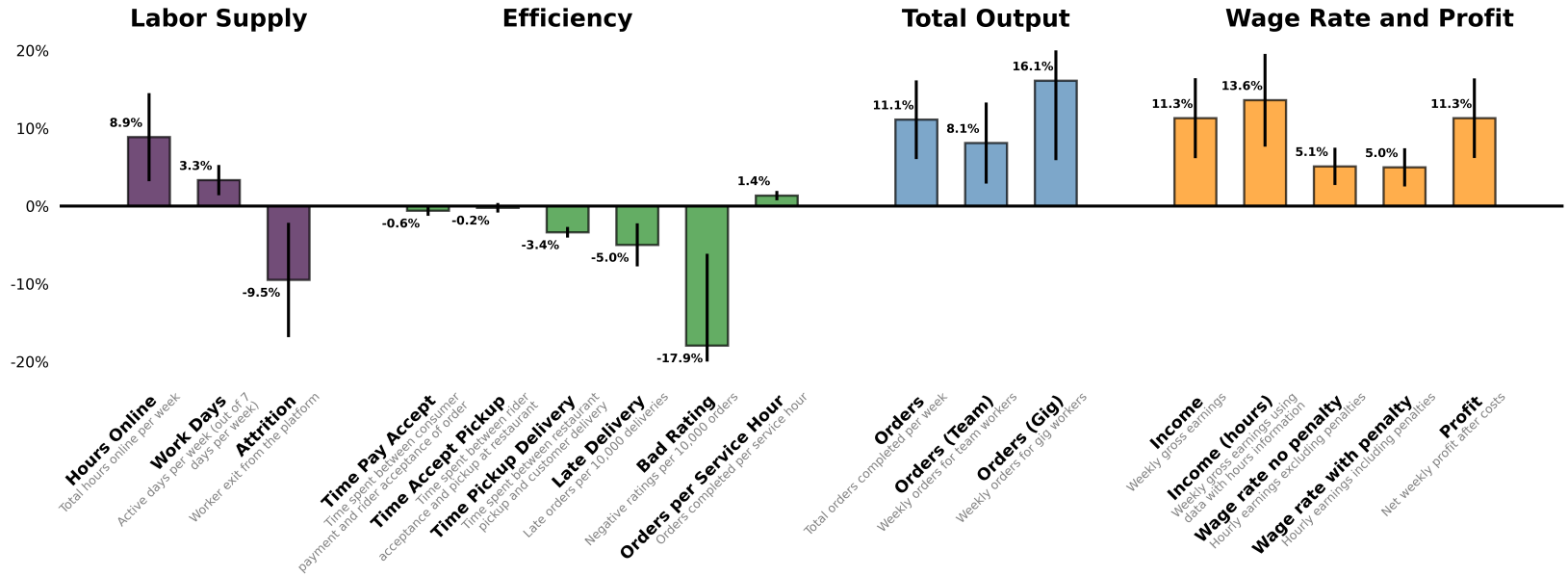
Notes: This figure replicates the analysis in Figure 3, but restricts the sample to drivers present Pre- and Post-AI, to account for the concerns for attrition.

Figure A.7: Threat to Identification: Restrict to Observations with Observed Online Duration



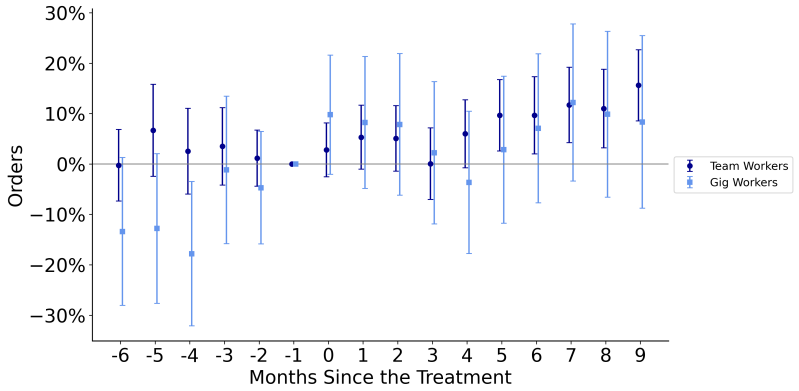
Notes: This figure replicates the analysis in Figure 3, but restricts the sample to observations with observed online duration.

Figure A.8: Threat to Identification: Treatment Effects with Additional Controls

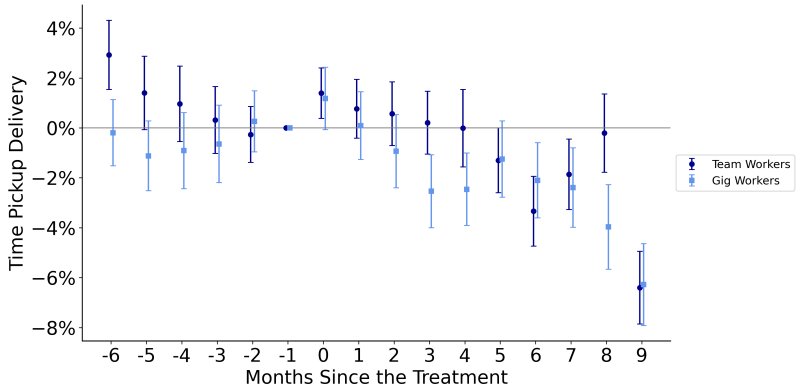


Notes: This figure replicates the analysis in Figure 3, but control for the difficulty of tasks, which is approximated by distance of the order.

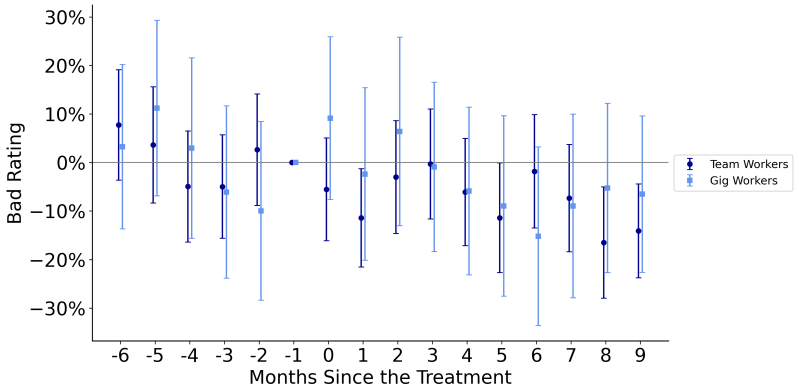
Figure A.9: Event Study Analysis: Productivity (Gig and Team Workers)



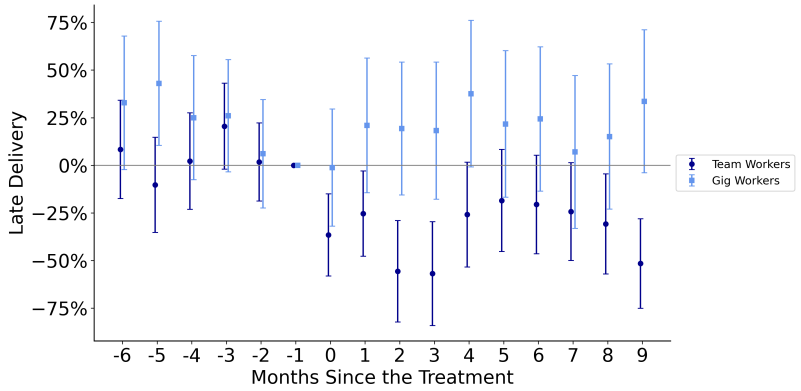
(a) Number of Orders Completed



(b) Time Spent Between Store → Delivery (Per Order)



(c) Bad Ratings



(d) Number of Delayed Orders

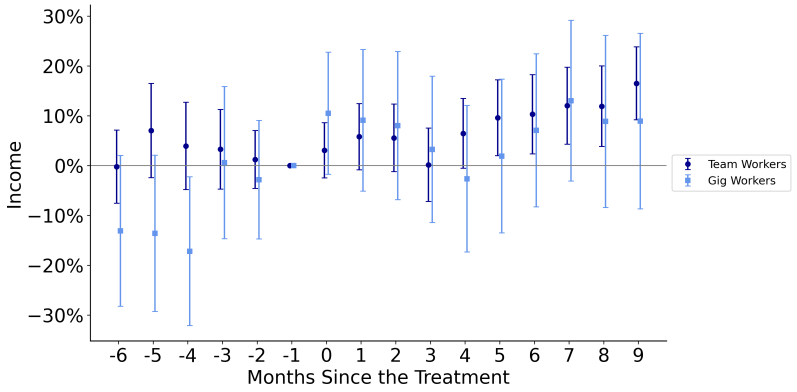
Notes: Figure A.9 presents an event study analysis of the introduction of the Intelligent Outbound Call (IOC) system, examining trends before and after the intervention for outcome measures related to productivity.

Panel (a) plots the difference in weekly orders between disabled and non-disabled workers for each time period relative to the IOC implementation. The x-axis represents the number of months relative to the intervention, with 0 indicating the implementation week, while the y-axis shows the estimated difference in orders between disabled and non-disabled workers.

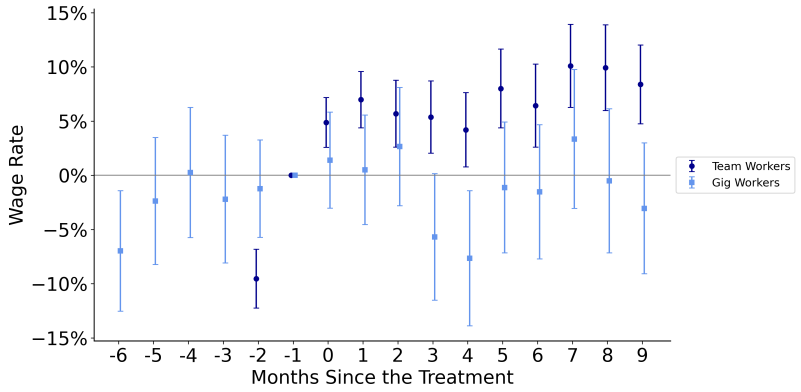
Panel (b) shows the gap in delivery time for the final stage (*Store → Customer*) between disabled and non-disabled workers, plotted relative to the implementation of IOC.

Panel (c) and (d) report differences in the incidence of *Negative Customer Ratings* and *Late Deliveries*, each normalized per 10,000 orders.

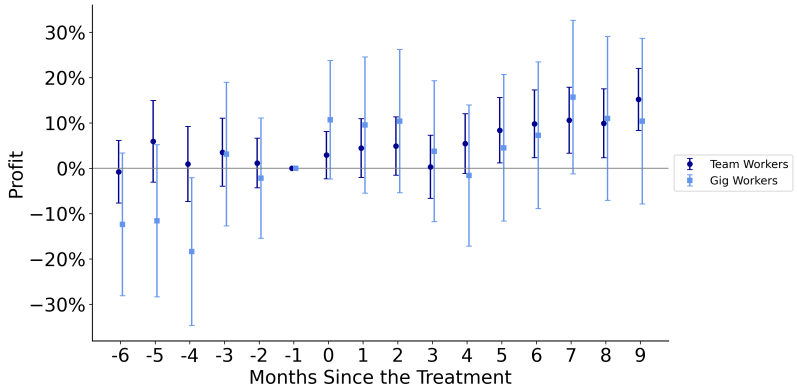
Figure A.10: Event Study Analysis: Wage and Profits



(a) Income



(b) Wage



(c) Profit

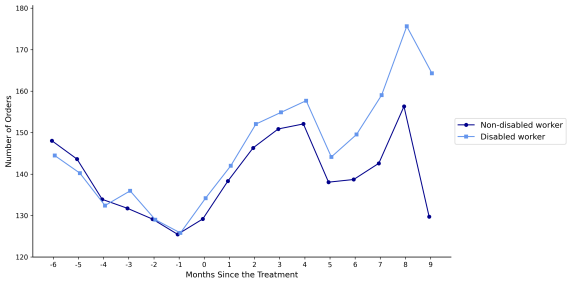
Notes: Figure A.10 presents an event study analysis of the introduction of the Intelligent Outbound Call (IOC) system, examining trends before and after the intervention for outcome measures related to wages and profits.

Panel (a) plots difference in income (measured by weekly gross earnings) between disabled and non-disabled workers across time periods relative to IOC implementation. The x-axis indicates months relative to the intervention, with 0 marking the implementation week, and the y-axis shows the estimated difference in gross income.

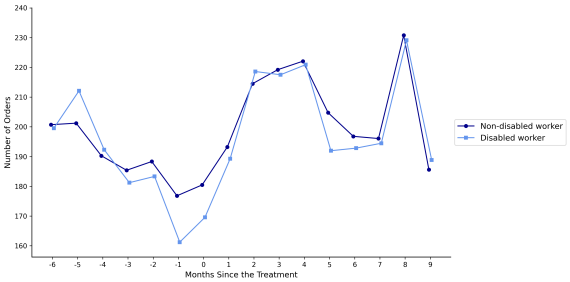
Panel (b) shows Difference in wage (measured by average earnings per hour online).

Panel (c) shows Difference in profit generated (measured by net weekly profit for the platform after deducting operating costs).

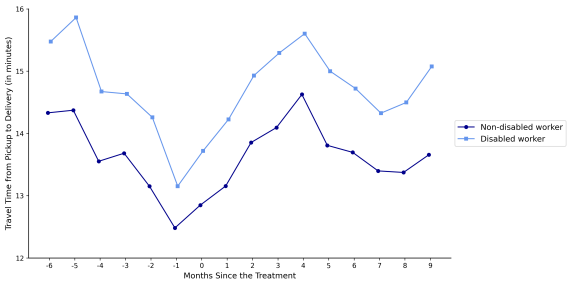
Figure A.11: Comparison of Raw Average Between Treatment and Control Groups



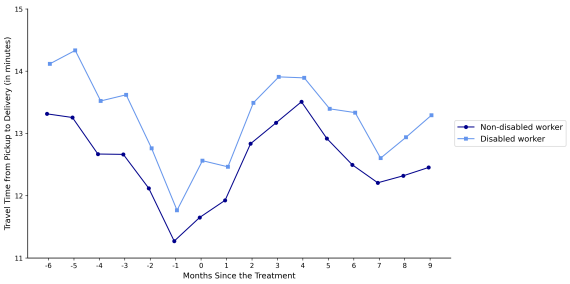
(a) Number of Orders Completed



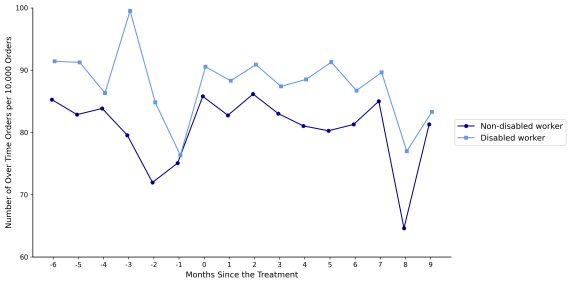
(b) Number of Orders Completed



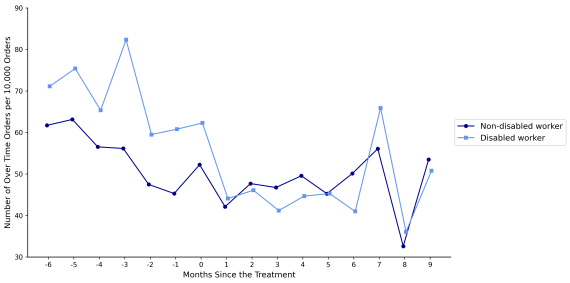
(c) Time Spent Between Store → Delivery



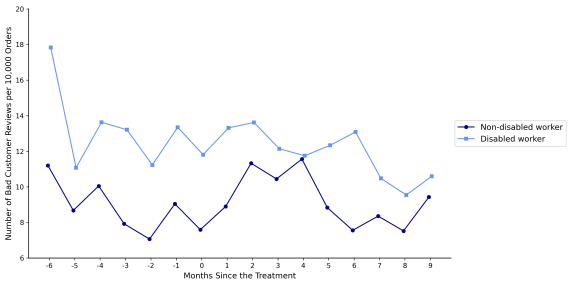
(d) Time Spent Between Store → Delivery



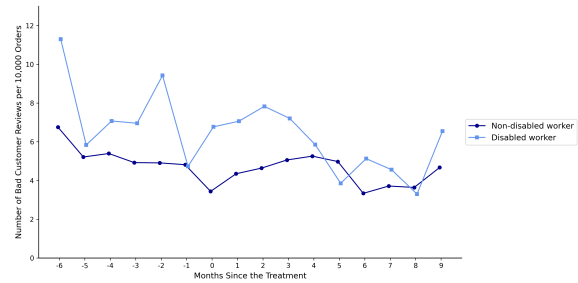
(e) Late Delivery Rate



(f) Late Delivery Rate



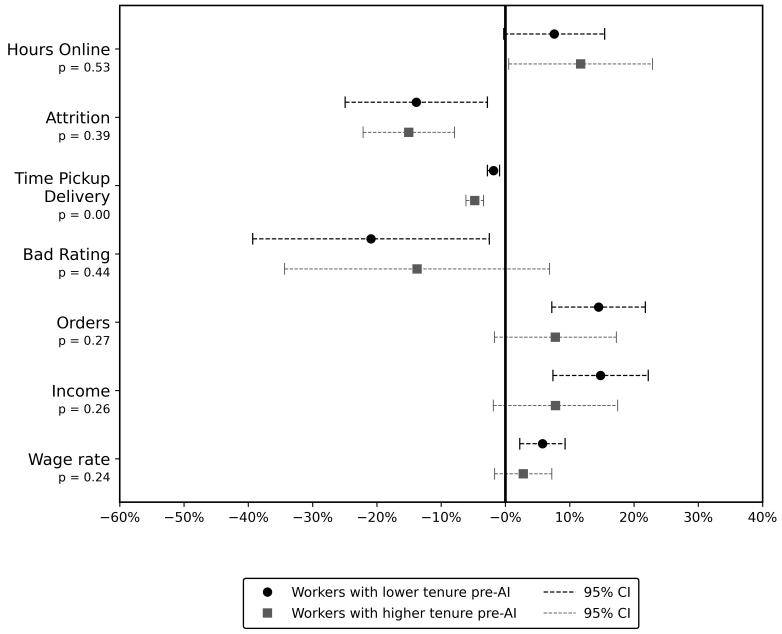
(g) Bad Rating



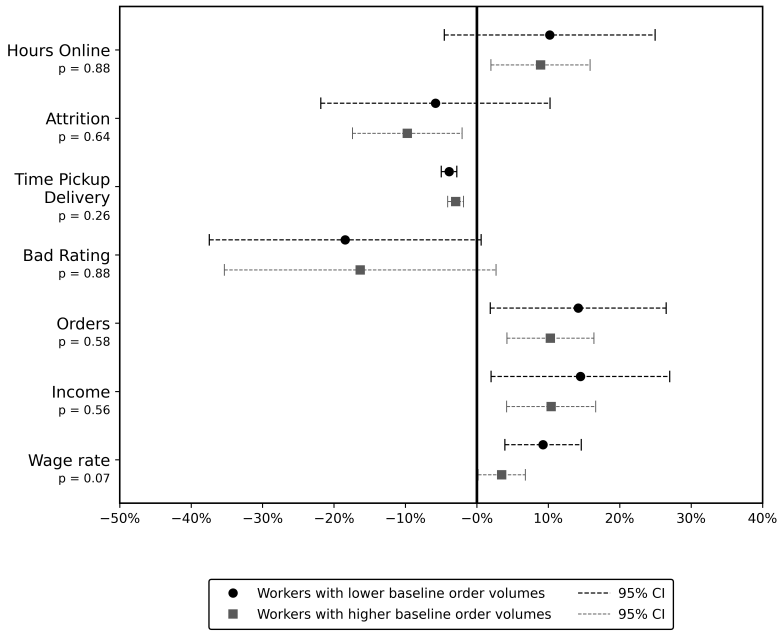
(h) Bad Rating

Notes: Here we compare the raw averages of the outcome variables between disabled workers (light blue) and non-disabled workers (dark blue). The left panel shows results for all workers, while the right panel shows results for those who remained continuously active on the platform (non-exit) throughout the study period.

Figure A.12: Heterogeneity by Pre-AI Tenure and Pre-AI Productivity



(a) Heterogeneity by Pre-AI Tenure



(b) Heterogeneity by Pre-AI Productivity

Notes: Figure A.12 presents the heterogeneous treatment effects by pre-AI tenure and pre-AI productivity. Pre-AI tenure is categorized as Low or High using a median split across workers, and pre-AI productivity is partitioned based on pre-AI order volumes, with the threshold set to the average weekly orders completed by the median worker.

B Entry of Disabled Workers

This section examines how the introduction of new technology affects the entry of disabled workers into the labor force. In late April 2022, the platform introduced the "Intelligent Outbound Call" feature, enabling disabled workers to initiate automated voice calls to customers directly through the app. Shortly afterward, in September 2022, a competing company launched a similar function. These two companies are the major players in this industry.

To analyze the impact of this technology on disabled workers' labor force participation, we begin with a cross-sectional dataset of workers. Each worker is classified as disabled or non-disabled, and their entry date is recorded. We then aggregate the data into disabled/non-disabled-week level observations, with each entry representing the number of workers. Using this dataset, we estimate the following difference-in-differences (DID) regression model:

$$Y_{ct} = \gamma \, \mathbb{1}[Disabled_c = 1] \times Post_t \; + \; \lambda_c \; + \; \xi_t \; + \; \varepsilon_{ct}. \tag{B.1}$$

where c denotes the disabled/non-disabled group, and t represents week fixed effects. Y_{ct} is the inverse hyperbolic number of new entrant in group c week t. We are interested in the estimates of γ.

Table B.1: Regression Results of Number of New Disabled Worker Entry

	(1)	(2)
Intelligent Outbound Call	0.261***	
(Focal Company Launch)	(0.110)	
Intelligent Outbound Call		0.193
(Only Focal Company Launch)		(0.190)
Intelligent Outbound Call		0.276**
(Competitor Launch)		(0.116)
Constant	4.894***	4.894***
	(0.0414)	(0.0415)
Observations	312	312
R-squared	0.924	0.924

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

Table B.1 presents the regression results. Following the introduction of the Intelligent Outbound Call feature, the platform experienced an average gain of 27.9% more disabled workers compared to the baseline level. This effect remains statistically and economically

significant even after the competing firm launched a comparable feature in September 2022, indicating robustness to rival technological adoption.

C Ethnographic Observations

Interviews

After conducting a preliminary analysis of the data on measuring the productivity of disabled workers, we sought to better understand their daily routines, how they work, and the challenges they encounter. Therefore, we conducted interviews with several disabled workers.

Figure C.1: Interviews with Disabled Workers



During these interviews, we spoke with seven disabled workers. To facilitate communication with hearing-impaired workers, we hired a translator proficient in sign language (She is seated at the bottom right of the picture, wearing a black shirt.).

Below, we summarize the demographics of our interviewees as well as the questions asked and the responses received.

Q: During pickup, do you encounter communication barriers?

A: Workers reported experiencing occasional communication barriers during pickups. Four respondents described needing to repeatedly remind restaurant staff—often two to three times—due to communication difficulties, while non-disabled riders typically resolve issues swiftly by shouting a single instruction that kitchen staff immediately understand.

Table C.1: Summary of Interviewee Demographics

Age	Hometown	Years of experience	Marital Status	Gender
22	Fujian Province	1+ years	Single	Male
26	Chongqing Municipality	3 years	Single	Male
27	Chongqing Municipality	4 years	Single	Male
24	Henan Province	3 months	Single	Male
25	Henan Province	3+ months	Single	Male
24	Henan Province	1+ years	Single	Male
23	Henan Province	1+ years	Single	Male

Consequently, non-disabled riders tend to receive meals faster because their vocal presence exerts greater urgency on kitchen staff.

Q: What specific challenges have you encountered during deliveries?
A: Workers highlighted several challenges:

1. One rider described an incident where a customer interaction required mediation by the station manager due to communication barriers. Despite the manager's assistance, the rider received a negative review, significantly affecting his morale and leading him to consider leaving the job.

2. Another rider expressed frustration with having multiple deliveries labeled as "T3" within one month, indicating these deliveries exceeded the app's allowed time window by at least three minutes. Such labels negatively impacted his performance record.

    - **Note on T3 classification:** A delivery is categorized as "T3" if it surpasses the app's specified time window. Although minor delays within a designated buffer period typically do not incur immediate penalties, significant delays adversely affect riders' service scores, potentially leading to reduced order allocations or other penalties. The expected delivery timeframe is calculated based on factors like the rider's speed, distance to the customer, and restaurant preparation time.

3. A rider mentioned taking proactive measures during deliveries to reduce confusion, such as sending a pre-written explanatory message to customers with the station's contact information to address issues caused by delays or miscommunications.

Q: Have you experienced forced stops during deliveries?

A: Four workers reported that switching between apps, such as moving from navigation to the order interface, frequently results in forced stops. One rider specifically noted that managing two phones simultaneously—one dedicated to navigation and another to orders—necessitated frequent stops, causing significant disruptions.

Q: Do you receive advance notifications about potentially overdue orders?

A: All workers acknowledged receiving in-app reminders, but due to the nature and pace of their work, they rarely had sufficient time to read or react to these notifications.

Q: Do you have access to rest facilities or water during work?

A: Workers reported using specialized restroom-locating maps to find nearby public facilities for quick rest and water. They seldom return to the station for breaks.

Fieldwork

To deepen our understanding of the working conditions and everyday challenges faced by disabled delivery workers, we undertook immersive fieldwork alongside them. Specifically, we conducted two practical explorations: First, one author simulated the experience of a newly registered worker by delivering multiple orders while wearing earplugs to mimic hearing impairment (see Panel (a) of Figure C.2). This allowed us to better understand the initial challenges faced by novice workers. Second, another author spent several days directly observing disabled workers by accompanying them full-time, seated on the back seat of an electric motorcycle during their delivery shifts (see Panel (b) of Figure C.2). This hands-on approach enabled us to closely observe and document the workers' daily routines and the specific difficulties they encounter.

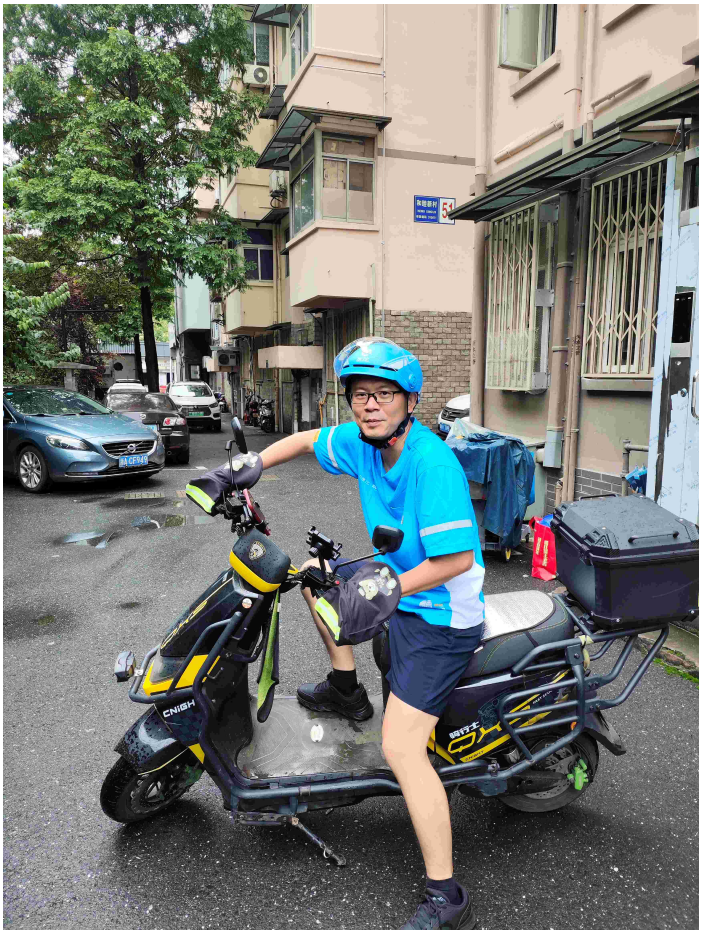
In the text that follows, the first person "I" refers to the perspective of a coauthor instead of a subject.

Fieldwork Records, July 2, 2024

Location: Hangzhou
Subject: Male, 27 years old, 3 years of employment
Duration: 10 hours 50 minutes (10:00 AM – 8:50 PM), 33 orders delivered
Observations:

1. Delivery workers tend to gather at a common meeting point after completing their orders. For instance, we were near Hangzhou 303 Life Plaza, where many workers

Figure C.2: Immersive Fieldwork with Delivery Workers



(a) Author delivering orders with earplugs to simulate hearing impairment



(b) Author observing a disabled worker during deliveries

frequently returned. Workers tend to avoid orders exceeding 4 kilometers, and after delivering such "distant orders," they usually navigate back to common gathering points.

2. A common tactic I noticed was workers pressing multiple elevator buttons when making deliveries. For instance, if a worker had to deliver to the 20th floor, he would also press buttons for the 21st and 22nd floors. This strategy allows them to reduce waiting time on the way down.

3. Over the course of a day, I noticed that many of the deliveries were from the same restaurants or to the same apartment complexes. The worker often made repeated trips to the same locations throughout the day.

4. During the lunchtime peak hours, the worker was handling six orders simultaneously, which led to a delay in one of the deliveries. After this delay, the number of orders assigned to them decreased significantly. One reason for the delay was that the second order was held up due to the restaurant's slow preparation and packaging process. The authors confirmed with the platform that no additional time is granted for subsequent orders the worker is simultaneously carrying in such situations. Another cause was the worker's decision to wait for the delayed order instead of collecting food for the other orders from nearby restaurants in the same mall. The worker could have optimized their pickup sequence to avoid unnecessary waiting.

5. The worker mentioned that if a delayed order is transferred to another worker, the worker who accepts the transferred order is not penalized for the lateness of the delivery. Workers are allowed to transfer up to two orders on the platform every 24 hours.

6. While having lunch together, I noticed that many restaurants offer discounts to the food-delivery workers, often around 20% off. After resuming work at 2 PM, the majority of the afternoon orders were for milk tea.

7. At 3:50 PM, the worker received a grocery order, after which he stopped to change the battery of his electric motorcycle. A fully charged battery allows for approximately 100 kilometers of travel. These batteries can be swapped at dedicated charging stations, which function like lockers, where you insert your empty battery and exchange it for a fully charged one.

8. At 4:30 PM, we returned to the gathering point in front of Plaza 303 to wait for more orders to be assigned. During this time, other workers at the site shared their

progress: one had completed 31 orders, another had finished 25, and a newer worker had completed only 7.

9. The top-performing worker at this station managed to complete approximately 2,823 orders in a month, earning around 20,000 CNY ($2,800 USD). However, this required working over 16 hours a day. At times, the worker would fall asleep on their electric motorcycle between orders, waking up only when the next one was assigned. Occasionally, they would even fall asleep while waiting at traffic lights.

Fieldwork Records, July 7, 2024

Location: Hangzhou
Subject: Male, 24 years old, 1 year of employment
Duration: 12 hours 20 minutes (7:45AM – 8:05PM), 75 orders delivered
Observations:

1. At 7:45 AM, we started the day. The worker provided me with a full set of sun protection gear, including a sun-protection suit and a neck gaiter. The temperature that day ranged from 38 to 40 degrees Celsius (over 100 degrees Fahrenheit). Figure C.4 shows the equipment required for working outdoors.

2. By 9:00 AM, we had completed eight orders. A fellow worker, mentioned in the previous report as the station's top performer, had already completed 21 orders. The worker informed me that the top performer starts his day at 6:30 AM and finishes at 2:00 AM.

3. Around 9 AM, we received an order for a large watermelon. However, the mall where the order was placed didn't open until 10:00 AM, so we had to pick it up from a locker in the underground parking lot. The security guard wouldn't allow us to ride the motorcycle down to the parking lot, so we had to run from the garage entrance to the basement, grab the watermelon, and run back up to avoid a late delivery.

4. Later in the morning, we received an order for McDonald's. The Platform indicated that the food should be picked up from a collection locker. When we entered the store, a staff member in the kitchen shouted, 'The food is on the counter,' but the worker didn't hear and went straight to the locker. After unsuccessfully entering the pick-up code several times, the worker finally went to the counter to communicate with the staff and retrieve the meal.

5. The intelligent outbound call is frequently used when workers request customers to open doors, and it is generally very effective. However, there was one exception during

our deliveries. In this case, when we arrived at the destination, the worker followed the customer's instructions to contact the apartment complex management via the intercom. After the system connected, no one spoke on the other side. Although I heard the door unlocking sound, the door did not automatically pop open, and there was no visible indication that it had unlocked. The worker continued waiting by the intercom until I eventually walked over and pulled the door open.

6. One order was delivered to an urban village, but the address was inaccurate. Upon arriving, the worker used the intelligent outbound call to contact the customer, who said, "I'm upstairs." However, it was unclear what the customer meant by 'upstairs.' I then heard the customer shouting from inside a factory behind us, but by that time, the worker had already started riding in the opposite direction. I stopped the worker, informing him that the customer was inside the factory, and we managed to find the customer. In such situations, a worker with a hearing disability could waste significant time locating the exact address. Later that afternoon, we encountered a similar issue with an order delivered to a construction site.

7. There was another incident where a restaurant incorrectly packed an order. The customer had ordered noodles and fresh tomatoes, but the restaurant forgot to include the tomatoes. The restaurant called the worker during delivery and asked them to return to pick up the missing tomatoes. However, we were already halfway to the destination, and returning would have delayed the delivery. Communication between the restaurant and the worker was challenging, even with the intelligent outbound calls. I had to step in to explain to the restaurant that we were near the destination and couldn't return. I suggested the restaurant to contact the customer directly to explain the situation.

8. In general, when a situation requires in-depth communication with restaurants or customers, a worker with a hearing disability still faces difficulties and takes significantly more time to resolve the issue compared to a non-disabled worker. While these types of problems are difficult to address, they occur rarely. In most cases, the intelligent outbound calls are sufficient to assist workers.

9. For long-distance deliveries, the worker drove at high speeds (40-60 km/h). Sitting in the back seat, the wind was so strong that it caused my eyes to tear up, and I could barely keep them open despite wearing glasses. I couldn't imagine how the worker managed to see clearly without any protective eyewear.

10. There were many heartwarming moments throughout the day. For example, building janitors and restaurant staff were noticeably warmer toward me (as I wore the food-

delivery worker uniform) compared to when I was a customer in my everyday life. They were more inclined to chat with us. During a short break inside a building in the afternoon to escape the heat, a janitor approached us, commented on how hot it was, and reminded us to stay hydrated. Later, during dinner, a restaurant staff member came over and spoke with us for a long time, discussing strategies to encourage restaurants to offer discounts to delivery workers.

11. It was clear that pedestrians, motor vehicles, and bikers were cautious—if not outright afraid—of us on the road, often stepping aside as soon as they saw our electric motorcycle approaching.

12. One moment that left a strong impression on me was when I asked the worker, "Are you happy working as a food-delivery worker? How do you feel?" The worker responded, "I have no emotions; I'm just here to deliver orders." He explained, "If you get emotional while delivering, you won't last long. Whether it's riding on the road, picking up orders, or delivering them— dealing with restaurant delays in preparing the food, climbing six or seven floors on foot, late deliveries, or carrying overweight packages—there's no room for emotions." Doesn't "having no emotions" mean there are many tough moments to endure? If everything were pleasant, would there be a need to become so desensitized? After a day of running deliveries, I felt the same — under the scorching sun, with temperatures over 38 degrees (100°F), sprinting with two boxes of milk and four bottles of water, climbing four flights of stairs, and racing from the 20th floor to the first in three minutes when an order was nearly late. In those moments, the feeling of happiness seems irrelevant. All you can do is toughen up and adapt to this lifestyle.

13. By evening, I was too exhausted to run. Many of the late-night orders required entering residential complexes on foot (because the security guard wouldn't allow us to ride the motorcycle inside) and climbing stairs, all of which the worker handled alone. After a full day of work, my fingers had pruned from being soaked in sweat for so long. All of this made me feel awkward, as the worker is much younger than I am but clearly much tougher, both physically (evidently) and perhaps mentally as well.

14. In the afternoon (before dinner), there were fewer orders, mostly for milk tea or desserts. One incident stood out that I want to note: we went to a dessert shop where only one young woman was preparing the orders, and it was clear she was overwhelmed by the number of orders. As three or four more workers arrived and waited for their orders, they began pressuring her to speed up. One worker was noticeably

impatient and even rude, saying, "If there's only one of you, what kind of restaurant do you think you can run? You're making us wait too long."

15. The worker I worked with mentioned, "I typically start running orders at 7:00 AM and take a break around 3:00 PM. Then I go back online from 4:00 PM until 8:00 or 9:00 PM." My observations and experiences working alongside different full-time workers were similar. According to them, to earn a decent income, they need to work between 10 and 13 hours a day.

16. Below is a summary of orders that I found particularly difficult for the workers to deliver:

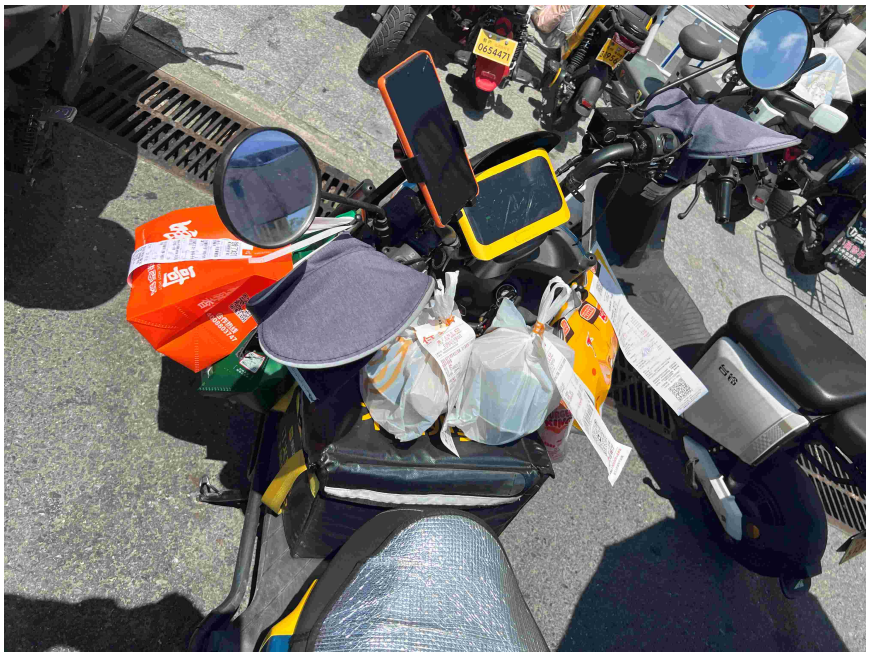
    - Some residential complexes have very slow elevators (there are many residents, but very few elevators).
    - Some older complexes do not have elevators.
    - Security guards at certain complexes do not allow workers to enter with electric motorcycles.
    - Apartments in large residential complexes can be difficult to navigate.
    - Delivering to urban villages is particularly challenging, as the locations are difficult to find and often require climbing stairs without elevators.

17. Below is a summary of some observations on how workers interact with the platform's algorithms:

    - Workers must perform three actions during the delivery process: 'confirm arrival at the store,' 'confirm food pickup,' and 'confirm delivery.' The app only allows workers to confirm arrival at store within a certain range of the store. However, if the restaurant has not prepared the food within three minutes after confirming arrival, the worker can report the delay, and the platform will grant an additional five minutes for delivery. To avoid potential delays when handling multiple orders, workers will sometimes confirm arrival at the store for order A, pick up order B from another restaurant, and then report a delay in food preparation for order A to gain more time. The risk is that false reports of delays can result in penalties, such as a lower service score (which affects priority in order assignment). I noticed that the worker I was paired with had multiple score deductions in his history due to false reports, though none occurred on the day I worked with him.

- To avoid being assigned too many unwanted orders by the station manager or platform, workers may delay clicking 'confirm delivery.' For example, when we delivered to Westlake University in the afternoon, the worker didn't want to take any more distant orders, so he delayed confirming the delivery.
- To return to a specific commercial area, such as Plaza 303, workers can accept orders in that direction from the order-grabbing page, and the algorithm will then assign more orders headed that way. There are two ways workers receive orders: they either get them through direct assignment (80% of orders are assigned this way), or if they want to take on more, they can choose orders from the order-grabbing page.
- Compared to the other workers I worked with, this worker was clearly more skilled at route planning and time management, as he avoided any late deliveries. He prioritized deliveries based on estimated times, rather than strictly following the algorithm's sequence, which helped him stay on time.

Figure C.3: Concurrent Orders During Peak Hours



Notes: Figure C.3 illustrates multiple orders being simultaneously handled by a delivery worker during peak hours. Experienced workers typically manage between 6 and 10 concurrent orders during peak hours. The maximum number of orders a rider can handle simultaneously is set by the platform's algorithm, considering factors like tenure and past performance. Further details on the assignment algorithm are discussed in Section 3.

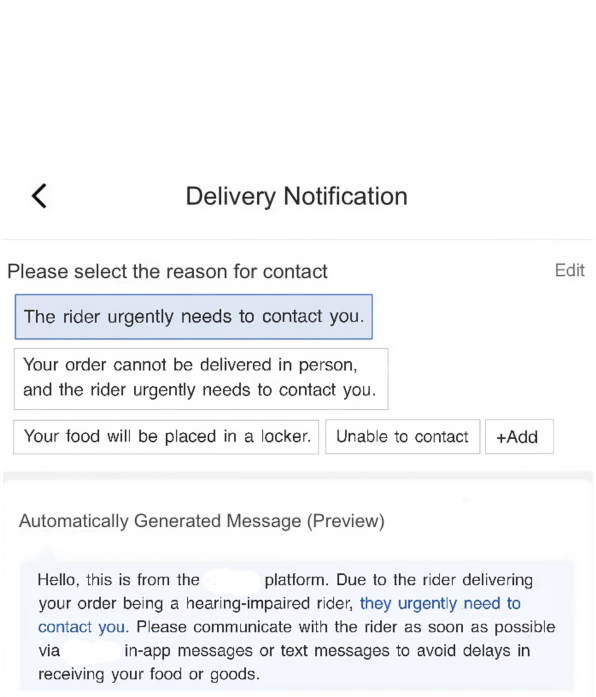
Figure C.4: Essential Equipment for Outdoor Work



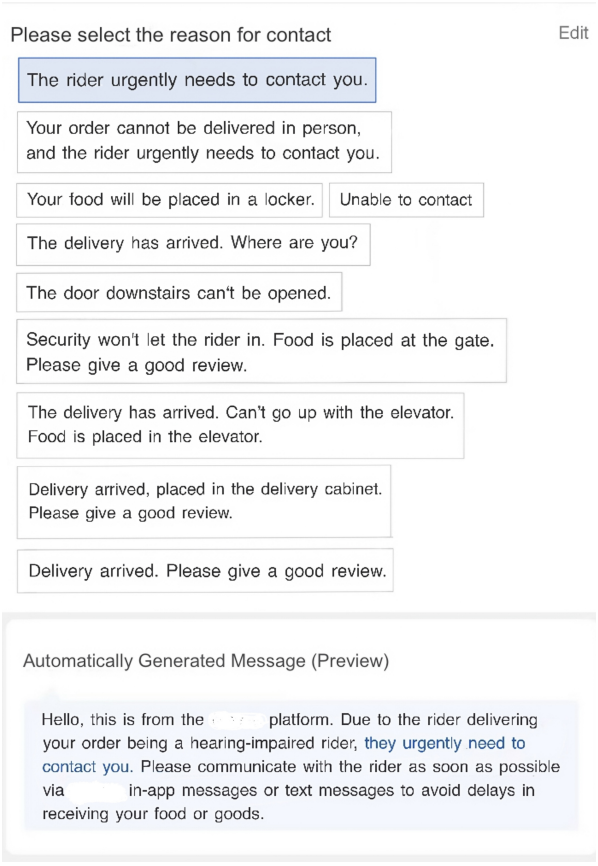
Notes: Figure 8 illustrates the necessary protective gear for outdoor delivery work, including a sun-protection suit, helmet, and neck gaiter. Typical summer temperatures in the focal city range from 38 to 40 degrees Celsius (over 100 degrees Fahrenheit). Consequently, complete sun-protection equipment is essential, and workers must remain well-hydrated to avoid heat-related illnesses such as heat radiation sickness.

D Intelligent Outbound Call

Figure D.1: Screenshot of the Intelligent Outbound Call Feature



(a) Default Prompts



(b) Customized Prompts

Notes: Figure D.1 shows the default prompts for Intelligent Outbound Call in Panel (a). Panel (b) presents a screenshot of one of the disabled workers we worked with, demonstrating the common prompts they use.

The Intelligent Outbound Call feature is an in-app tool designed to empower delivery workers with hearing or speech impairments by facilitating effective communication with customers through automated voice calls. Previously, these workers relied solely on text messaging within the app, which customers often overlooked, leading to misunderstandings and delays. This technology bridges a critical communication gap by converting selected text prompts into automated voice calls.

The Intelligent Outbound Call feature is integrated into the delivery app. Workers can select from predefined default prompts or input customized prompts relevant to specific situations. These text messages are converted into audio through an advanced text-to-speech system and sent to the customer via a cellphone call directly from the worker's phone number.

Panel (a) of Figure D.1 displays default prompts provided by the app, addressing common scenarios such as urgent contact needs, inaccessible delivery locations, or delivery placement instructions. Workers can quickly select an appropriate option, preview the automatically generated message, and initiate the automated call instantly.

For situations beyond standard prompts, workers have the flexibility to create and store personalized messages. Panel (b) of Figure D.1 illustrates common customized messages tailored to real-world delivery challenges, such as access restrictions, security issues, or polite customer engagement.

The Intelligent Outbound Call feature significantly enhances the delivery experience in three key ways:

- Improved Customer Responsiveness — Customers are more likely to notice and respond to automated voice calls compared to in-app text notifications, leading to faster resolutions and fewer misunderstandings.

- Enhanced Efficiency — By addressing communication bottlenecks through automated messages, workers with disabilities can complete more deliveries with fewer delays, significantly improving productivity.

- Flexible and Personalized Communication — Workers can easily adapt their messages to specific circumstances, clearly communicate unique instructions, or politely request customer assistance or feedback, thereby enhancing the overall delivery service quality.

In essence, Intelligent Outbound Call technology represents a significant advancement in inclusive communication practices within delivery services. It empowers workers with disabilities to operate efficiently and confidently while ensuring high-quality service for customers.