Empowering Inclusive Work*

Yanyou Chen
University of Toronto

Mitchell Hoffman
UC Santa Barbara and NBER

Huilan Xu
Zhejiang University

Zhe Yuan
Zhejiang University

January 2026

Abstract

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

Keywords: AI, productivity, disability, personnel economics, gig work
JEL Codes: M50, J14

*We thank David Autor, Heski Bar-Isaac, Gordon Dahl, Chiara Farronato, Peter Kuhn, Xiang Hui, Rafael Lalive, Sarah Miller, and Imke Reimers for helpful comments, as well as seminar participants at Beijing, Bocconi, Cornell, CUHK, EIEF, Gothenburg, Guelph, Lausanne, NBER Labor Studies, Shanghai, Toronto, Toronto Rotman, UCR, UCSB, Wisconsin, and Zurich. We also thank Xiaonan Zhu, Weiyue Zhang, Ruibo Sun, Yaqi Du, and Yibing Xia for excellent research assistance. We thank Schmidt Sciences for their support. Any opinions, findings, conclusions and recommendations expressed are those of the authors and do not necessarily represent the views of the focal Platform.

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[1] 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 firms 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, labor supply, output, wages, and profit for firms?

2. Can AI tools affect these outcomes for 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 millions of active workers and completing hundreds of millions of deliveries per month. Second, the Platform follows an open hiring policy: all applicants who meet basic criteria are accepted, with no differential standards 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

[1]While some use the *person-first* "worker with a disability," many disability scholars prefer the *identity-first* term "disabled worker" (Gernsbacher, 2017; Grech et al., 2024). We use both, but often the latter.

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

We focus on workers who are deaf or hard of hearing (DHH), who constitute the majority of disabled workers on the Platform, and compare their outcomes to those of non-disabled workers.[2] 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.

As noted in Section 1, we use a rich administrative dataset from the Platform, which includes all newly registered DHH 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 workplace outcomes of DHH workers to non-disabled workers on many dimensions. The data include worker productivity variables capturing both the quantity of work, such as speed at different stages of the delivery of work, and work quality, as captured by customer ratings. The data further include labor supply variables, on both the intensive and extensive margins, as well as information pertaining to worker compensation and Platform profits.Crucially, in the middle of our observation period in April 2022, the Platform introduced a novel AI-based "Intelligent Outbound Calling" feature designed specifically to help DHH workers communicate with customers. We propose a simple conceptual framework about differences in workplace outcomes between disabled and non-disabled workers, and about how those differences are affected by AI tools.

Prior to the new AI tool, DHH workers work slightly slower than non-disabled workers (Section 2). This is especially so on the final segment of trips, from restaurant pickup to customer delivery, where communication is most important, where disabled workers are 4% slower. DHH workers are substantially more likely to experience "left tail" outcomes, being 8% more likely to have significantly late deliveries and 37% more likely to receive a bad customer rating in a given week (with the caveat that bad customer ratings are rare). Despite this, pre-AI, DHH workers have substantially higher labor supply: they work 7% more hours per week and are 32% less likely to quit. The higher labor supply of DHH workers more than compensated for their small efficiency gap in terms of total output, with DHH workers performing more orders per week, and actually achieving higher profits per week

[2]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.

than non-disabled workers. However, because DHH workers work slower, they earn 8% less per hour than non-disabled workers, an earnings gap in line with disability gaps in other settings.

As detailed in Section 3, the AI tool we study is a text-to-speech phone calling system offering preset and customized audio prompts, and that uses AI to make the calls sound natural. Rolled out suddenly to all DHH workers (but not non-disabled workers), we exploit this as a "natural experiment" to estimate via difference-in-differences how the AI tool affects workplace outcomes for disabled workers. The AI tool causes DHH worker performance to substantially improve. DHH workers' speed of work increases, especially on the trip segments involving communication where they lagged the most previously. They also have far fewer late deliveries and bad ratings, causing the pre-AI gaps between DHH and non-disabled workers to shrink by half. Orders per week increases, especially among DHH workers who work fully independently without the support of other workers. Labor supply and retention further increase. These improvements translate into a higher economic value of DHH workers, with Platform gains far larger than costs. AI causes the disabled worker pay gap to fall from 8% to 3%, a decrease of over half. While less severely impaired DHH workers benefit too, gain are largest among severely impaired DHH workers. Overall, the results support the key mechanism of the AI tool improving communication with customers (as opposed to alternative stories like the AI tool increasing customer sympathies for DHH workers).

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.[3] An important exception is Bloom et al. (2025) who find that the large rise in work from home has increased the employment of disabled workers. We are not aware of prior work in any academic field that compares disabled and non-disabled workers on a large scale in terms of individual productivity data.[4] The intervention we study, a simple non-generative AI tool, has the largest impact on closing

[3] 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.

[4] This reflects data constraints rather than lack of broad scholarly interest in understanding disabled worker productivity and challenges. Human–computer interaction scholarship uses interviews to examine challenges faced by DHH workers on gig platforms, including Uber (Lee et al., 2019) and food delivery in China (Chen et al., 2024a). The Uber study is based on interviews with five drivers, supplemented by qualitative analysis of in-app feedback, while the China study relies on interviews with six couriers and four station managers. Neither analyzes large-scale administrative data nor examines the impact of AI tools.

the disability wage gap of any intervention that we are aware of.[5]

Second, it contributes to work on the economics of AI. While AI use is accelerating, causal evidence on its impact for disadvantaged groups is scarce. Work examines 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 affects workplace outcomes for disabled workers. In doing so, it provides evidence of how AI can help reduce inequality (Autor, 2024). Our analysis of heterogeneous treatment effects suggests that disability may be critically different from other dimensions of workplace inequality like lower productivity or lower tenure, as we find heterogeneous effects only by severity of disability. While generally not based on administrative data and usually lacking quasi-experimental variation in technology, research on personal computers suggests some of role of technology in improving outcomes for disabled workers (Kruse et al., 1996; Krueger and Kruse, 1995). It also connects to the broader economics literature on IT and productivity in firms (Bresnahan et al., 2002; Bartel et al., 2007).

Third, our paper contributes to work in personnel and organizational economics on demand for and returns to disadvantaged workers. While there is strong interest in studying hiring returns in general (Oyer and Schaefer, 2011), there has been relatively scant work on firm returns from disadvantaged workers (Hoffman and Stanton, 2025). One recent exception is Benson et al. (2024), who analyze productivity differences between white and minority workers 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 to examine the *profitability* of disabled workers for an organization or platform. Pre-AI, while DHH workers are less efficient than non-disabled workers, they are still more profitable overall. This highlights the importance of capturing a broad set of outcome variables in evaluating disabled worker profitability. An analyst who focused only on efficiency variables, and ignored the labor supply variables of hours and retention, would draw mistaken conclusions about the relative profitability of DHH workers, a point we return to in the Conclusion (Section 4).

1 Context, Conceptual Framework, and Data

The Platform is one of the two leading companies in China's massive online food-delivery market. It operates similarly to services like Uber Eats or DoorDash but on a much larger scale. Together, the two main Chinese platforms engage approximately 10 million delivery

[5]Disability wage gaps seem quite stubborn.

workers at a time, making this a labor force larger than that of some countries.

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 Platform is available nationwide, with a higher concentration of workers in urban and eastern regions of China where food delivery is most prevalent.

Our engagement with the Platform. Our engagement with the Platform on this project began in late 2023. One of the coauthors, who had received an order from a DHH worker, reached out to the Platform about the possibility of evaluating the impact of the AI tool. While the Platform frequently evaluates the impacts of changes to new customer features or other app improvements, it had never evaluated the impact of the AI tool for DHH workers, but seemed keen to better understand its impact.[6]

"Hiring" policy. The Platform maintains an open registration policy (referred to hereafter as "hiring," though drivers are independent contractors). To begin working, an applicant must meet three universal requirements: be at least 18 years of age, possess a valid driver's license, and have no disqualifying criminal record. Consequently, these standards are uniform across all applicants, with no separate criteria for workers with disabilities.

Disability status on the Platform. Across both major food delivery platforms, most disabled workers are DHH (*wú shēng* in Mandarin). The Platform defines a worker as being disabled based on the driver uploading their *certificate of disability* during registration. The certificate is a formal, government-issued document requiring medical evaluation by doctors and is used to determine disability insurance benefits. The Appendix provides a screenshot of the disability question during the registration process.

During registration, both pre- and post-AI, drivers are asked (1) what type of disability, if any, they have, (2) the disability level, (3) whether they consent to the Platform sharing their disability with customers so that there is more understanding in unexpected situations. Disability level for DHH workers is coded based on the Chinese national four-grade system (GB/T 26341-2010), where 1 denotes profound, 2 severe, 3 moderate, and 4 mild hearing disability. Drivers know their level since it is listed on their certificate of disability. In our data, we don't observe an answer to (3), but our understanding is that virtually all disabled workers select Yes.

Both pre-AI and post-AI, customers receive an in-app notification when their driver is DHH. However, many Chinese customers do not see notifications on their phone, reflecting that they get lots of notifications and can't attend to them all. This is a challenge, as there

[6]The Platform's driver operations team that we collaborate with does much less policy evaluation than other teams within the Platform.

is a general social expectation in China that customers will get a call from their driver when the food arrives.

Work models: team vs. gig. Workers on the Platform can choose between two work models: a *team* model, in which workers are assigned to stations and are paid under an accelerating piece-rate schedule, with unit prices that increase stepwise in the number of monthly completed orders, and a *gig* model, in which workers operate independently and are paid per delivery. Exact pay is determined algorithmically and varies across deliveries, with longer deliveries typically paying more. The team model is intended to be full-time while still allowing workers to choose their hours, whereas work in the gig model can be part time. Both gig and team workers are included in our data (about two-thirds are gig), and we control for work model in our analysis.

Matching of workers to deliveries. Gig and team workers also differ in order assignment. Gig workers primarily rely on independently selecting orders from the order pool (similar to UberEats and DoorDash drivers in the US).[7] 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.

The orders that appear for gig workers in the order pool, and the order that a team driver is assigned, depend on a driver's location, demand conditions, and the driver's contemporaneous *matching score*. The exact matching score is proprietary, and depends on a driver's (1) service score, (2) "Rider Level," and (3) order acceptance rate. The service score depends on customer reviews and late deliveries. The Rider Level is calculated based on tenure and the cumulative number of completed orders. Refusing system-assigned orders directly reduces a worker's Acceptance Rate. Workers report that this factor has the greatest impact on assignment score, so they usually do not reject system-assigned tasks. Critical, a worker's disability status does not affect their matching score.[8]

In sum, gig workers generally choose their orders and team workers are assigned orders. Both the order pool for gig and the assignments for team depend on workplace outcomes.

[7] While gig do receive some 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.

[8] More than assignments, the main thing affected by the matching score is the number of orders a driver is allowed to carry. The max 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.

1.1 Conceptual Framework

We present a simple framework 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.

Workers have a productivity or efficiency 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 zh, and workers get paid a simple piece rate w.

The return to work off the Platform is $\frac{\beta}{\beta-1}(1-h)^{\frac{\beta-1}{\beta}}$, where r measures the value of one's outside option and $\beta > 1$. The parameter, β, captures how sharply the worker substitutes between work on and off the Platform. Worker utility is thus given by:

$$U = w \cdot zh + r \cdot \frac{\beta}{\beta-1}(1-h)^{\frac{\beta-1}{\beta}}$$

In our empirical analysis, we measure productivity broadly, incorporating speed at different points, customer reviews, and average total time per order. The firm makes no decisions and earns profits $A \cdot zh$, where A is a piece rate measuring the Platform's commission.

Workers make a decision about the share of their work time to spend on the Platform. Using the first order condition of $\frac{\partial U}{\partial h} = 0$, we arrive at a simple analytical expression $h = 1 - r^{\beta}(wz)^{-\beta}$. Taking partial derivatives yields $\frac{\partial h}{\partial r} = -\beta r^{\beta-1}(wz)^{-\beta} < 0$ and $\frac{\partial h}{\partial z} = \beta r^{\beta} w (wz)^{-(1+\beta)} > 0$, i.e., people work more on the Platform when they have worse outside options and higher Platform efficiency.

Based on the literature and our ethnographic observations, we posit two main differences between disabled (denoted by D) and non-disabled workers (denoted by N):

Assumption 1. Disabled workers have worse outside options than non-disabled: $r_D < r_N$.

This assumption (A1) is consistent with a large literature, cited above, on the challenges and discrimination faced by disabled workers.

Assumption 2. In the absence of AI tools, disabled workers have lower efficiency: $z_D < z_N$.

This assumption (A2) is consistent what we gleaned from our conversations with disabled workers about work before the AI tool, namely, that there were often significant bottlenecks toward quickly making deliveries.[9]

[9]Disability scholarship stresses that differences between disabled and non-disabled individuals reflect not only differences between people, but also environmental features (e.g., discrimination, accommodations) (Krieger, 2010). Thus, A2 is not necessarily a statement about inherent capabilities, but rather a joint statement about capabilities and the environment.

Outcomes on the Platform can be grouped into those related to efficiency (z), labor supply (h), and total output (zh). The base framework and the assumptions yield four implications (I1-I4) that we examine using the data.

Implication 1. Before the new AI tool, disabled workers have lower efficiency than non-disabled workers.

Implication 2. Before the new AI tool, differences in labor supply between disabled and non-disabled workers are ambiguous. Disabled workers will work more on the Platform if the difference in outside options between disabled and non-disabled workers is large relative to differences in efficiency.

Implication 3. Before the new AI tool, average differences in total output and profits between disabled and non-disabled workers are ambiguous.

Implication 4. Suppose an AI tool increases the efficiency of disabled workers. This will increase their labor supply, total output, and profits on the Platform.

I1 is merely a re-statement of Assumption 2. For I2, the difference in labor supply is ambiguous because Assumptions 1 and 2 push in opposite directions: A1 makes disabled workers work more and A2 makes disabled workers work less. For I3, total output is zh, and since h is ambiguous, so is zh. For I4, by increasing z_D, an AI tool also encourages disabled workers to increase labor supply, and thus two forces together work toward increasing total output and profits.

This framework is deliberately simple and abstracts from many aspects of reality. First, there are many dimensions of worker productivity, including speed at different points, customer ratings, and late deliveries, while our model contains a single z. Second, our model abstracts from separate modeling of employee attrition—in our context, we think it is conceptually useful to think of "labor supply" as encompassing hours on the Platform, hours at various times of the day, and whether someone stays with the Platform.

1.2 Data

We obtained a proprietary administrative panel from the Platform, organized at the worker-week level and containing detailed measures of worker activity and performance. The panel covers workers who registered on the platform between 2021 and 2023 and completed at least one week of work, and includes all 4,284 workers who report DHH as their only disability, as well as nearly 40,000 non-disabled workers randomly sampled from the same population. In total, the data contain hundreds of thousands of worker-week observations.

The panel 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 DHH workers. We have over one year of pre-AI data to compare the performance of DHH and non-disabled workers, and we can track workers for more than a year to study the AI tool's impact. All variables are available for the full 2021-2023 period unless otherwise stated.

We examine a wide range of worker outcome variables, which we divide into four groups:

Efficiency: z in our framework. Efficiency or productivity refers to the speed and quality of work. Beyond overall speed, our data enable us to look at time spent during the three steps of the work process. 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.[10]

The three steps are as follows: First, once an order is initiated by the customer, it is assigned to a worker (Order $\rightarrow$ Worker). Second, once the worker receives the order, they pick it up from the restaurant (Worker $\rightarrow$ 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 $\rightarrow$ Delivery). Beyond overall speed, we also observe Platform-defined late delivery rate, which refers to deliveries that are late beyond the delivery window, indicating a significantly late order.

Quality is measured using customer ratings. Rather than use 1–5 stars as on Uber, customers on our Platform click a button to indicate a Bad Rating. Bad ratings are at the discretion of the customer and occur for different reasons, including late deliveries, not receiving the delivery, and food damage during delivery (e.g., spilled soup). Since bad ratings are highly skewed at the driver-week level (i.e., zero for most driver-weeks and positive for others), our main functional form for analyzing bad ratings is a dummy for whether a driver has any bad ratings in a week. In the Appendix, we show that our conclusions are unchanged when we instead use a count of bad ratings in a week, estimated using either OLS or count models. Drivers pay a penalty to the Platform for each bad rating received, giving these ratings direct economic consequences.

Accident data is not comprehensively collected by the Platform and is not provided to us.

Labor supply: h in our framework. Our two main ways of measuring labor supply are (1) a worker's weekly hours spent online (i.e., available for orders) and (2) whether the

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

worker quits the Platform. Beyond total hours, we also analyze days worked per week (i.e., so we can see if hours effects are coming from more days or more hours per day).

On (1), hours data are measured based on the worker logging into the app or the app is working in the background. Hours data only become available in our dataset starting in Week 19 of 2021, and only become available with a low rate of missingness starting in Week 9 of 2022.[11] The Appendix shows that our conclusions are robust to restricting to observations with non-missing hours data.

On (2), the data does not contain a formal flag for attrition. Instead, we proxy whether someone attrites based on having 0 weekly orders for 8 weeks in a row.[12]

Total output: zh in our model. This is measured primarily by a driver's orders completed per week. To shine light on where order effects manifest and task difficulty, we will also secondarily look at the total number of unique restaurants and customers served.

Income, wages, and profits. The administrative data we received does not contain fields for driver income, driver wages, and Platform profits. Instead, we use parameter information provided by the Platform, combined with our administrative data, to estimate these variables for each driver-week.

As on US delivery platforms, drivers are generally paid a piece rate per order. While the true piece rate varies by order, we estimate income using a single piece rate for gig workers and accounting for the accelerated piece rate profile of team workers. A driver's hourly wage is based on income divided by hours online. We calculate hourly wages two ways: (i) based on income alone and (ii) reflecting that drivers pay penalties when they receive bad ratings.

Platform profits from a driver in a given week, π_{it}, are equal to delivery fees from that driver's orders, minus the cost of any onboarding, minus the cost of any bad ratings:

$$\pi_{it} = \pi_0 y_{it} - c \cdot 1(\text{t is driver i's first week}) - k \cdot \text{Bad Ratings}_{it} \tag{1}$$

Profit per order, π_0, is a confidential number based on Platform records. Second, the recruiting and onboarding cost of a new worker, c, is assumed based on our discussions with the Platform to be around 100 CNY. The presence of this term accounts for any value DHH workers provide by having higher retention. Third, k is the Platform's cost of a bad customer rating. Based on our discussions with the Platform, we assign k to be 100 CNY, with k intending to capture the full costs of negative reviews, including effects on customer

[11] Before week 19 of 2021, hours is always missing. From week 19 of 2021 to Week 9 of 2022, hours is missing in 70% of weeks. Beginning in week 10 of 2022 and for the rest of the data, hours is missing in 6.8% of weeks.

[12] Our conclusions are similar under other reasonable definitions, including 0 weekly orders for 7 weeks in a row or 0 weekly orders for 9 weeks in a row.

retention and reputational spillovers, net of the penalty paid by drivers to the Platform.

Control variables. The data also include rich control variables about workers and the type of work performed.

Qualitative data. In addition to quantitative analysis, we conduct qualitative fieldwork to better understand the experiences of DHH workers, building on this practice in other recent studies of workers in organizations (Blader et al., 2020; Friebel et al., 2025). Performed in summer 2023, this includes:

1. Interviews by coauthors with 10 DHH workers to learn about their daily routines, challenges, and communication methods.

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

These qualitative insights help us interpret data patterns and to identify barriers faced by DHH workers, especially in customer communication, navigation, and service speed. They also provide context for understanding labor supply differences, as many DHH workers expressed a strong attachment to their jobs due to limited alternative employment opportunities. Appendix C provides further information on the qualitative data. We refer to qualitative findings where relevant when discussing the quantitative results.

1.3 Sample and Summary Statistics

Our analysis sample is based on excluding weeks where drivers have 0 orders, as these are weeks where drivers are not working.

Summary statistics. Table 1 presents summary statistics for our 43,605 food-delivery workers. Compared to non-disabled workers, DHH workers are slightly older (mean age 33 vs. 30 for non-disabled) and more likely to be male (93% vs. 91%), but are similar in work models: among both DHH and non-disabled workers, about 35-36% are team workers.

Panel C show the distribution of disability severity among DHH workers, showing a high concentration or 54% in the most severe category (Level 1).[13]

[13]In contrast to our sample where over 53% of DHH workers fall into the most severe category (Level 1), the general population of DHH individuals 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

Table 2 provides a detailed summary of all outcome variables. In the Appendix, we also similar tables separately for gig and team workers.

On average, workers complete about 150 deliveries per week, with a median of 127. 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 91 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 (212) compared to gig workers (73), reflecting the full-time nature of team-based work.

Orders take on average about 30 minutes for drivers in our sample. The time from order placed to work acceptance averages 3.4 minutes. The worker-to-store travel time averages 12 minutes, with a median of 11.5 minutes. The store-to-delivery time is slightly longer at 14 minutes, with a standard deviation of 5.1, with variability due to delivery distances and potential communication delays with customers. Per 10,000 orders, the late delivery rate is 416, meaning 4% of orders are substantially delayed.

Bad ratings, our measure of customer satisfaction, are much rarer. The bad rating rate per 10,000 orders is only 9.62, meaning only 1 in a 1000 orders yields a bad rating. Across driver-weeks, the median bad rating share is 0, i.e., most driver-weeks don't have any bad ratings.

Workers spend an average of 40 hours online per week. This is much higher than hours on Uber, where many driver work part time (Cook et al., 2021). The high standard deviation of 30 hours indicates wide dispersion across drivers in hours. The median number of workdays per week is 6, though the average is lower at 5.3, suggesting that while many workers work full weeks, others participate part-time or intermittently. Average hours are 58 for team workers and 23 for gig workers, as seen in Appendix Table A.1.[14]

The weekly attrition rate on the Platform is 8% for the new drivers we study, meaning the average driver lasts for around 3 months. While such an attrition rate is high, it is comparable to that in other low-skill jobs, e.g., the call-center worker studied in Burks et al. (2015). The exit hazard decreases quickly, so there is still is a sizable share of workers who stay much longer.

discrepancy suggests that our sample is skewed toward those with the most profound hearing loss. A natural explanation for this difference is simply that there may be less discrimination and lower barriers in gig work for people who are severely DHH (e.g., don't have to do an interview), as opposed to less likely to declare their disability to the Platform.

[14]The high intensity of platform work in China is a notable departure from patterns observed in the U.S. While the average Uber driver in the U.S. works approximately 15 hours per week (Chen et al., 2019), our data align with broader trends in middle-income countries where gig work often serves as a primary, full-time source of income. The 5.3-day average we observe is consistent with recent evidence on Chinese ride-hailing drivers, who typically work roughly 7 hours per day (Chen et al., 2024b).

The Appendix also presents summary statistics separately for gig and team workers. There is less variability within each work model separately, and we also present our main results separately by work type.

2 Pre-AI Differences by DHH Status

Our analysis proceeds in two stages. First, in Section 2, we measure differences in outcomes between DHH and non-disabled workers prior to the AI tool. Second, in Section 3, we evaluate the impact of the AI tool on outcomes of DHH workers (versus the comparison group of non-disabled workers) using a difference-in-differences approach.

2.1 Empirical Strategy

To compare outcomes between DHH and non-disabled workers in the pre-AI period, we regress each outcome on a DHH indicator and controls:

$$Y_{it} = \alpha + \beta \cdot DHH_i + \mathbf{X}'_{it}\gamma + \varepsilon_{it}, \tag{2}$$

where Y_{it} is the outcome for worker i in week t; DHH_i is a dummy being DHH; $\mathbf{X}_{it}$ are controls; and ε_{it} is an error. Throughout the paper, standard errors are clustered by worker, as DHH status varies by worker.

Our controls are relatively standard, and include year-week dummies, city of work dummies, age dummies, gender, team status (gig or team), and registration year dummies. In our main specifications, we do not control for average distance between the restaurant and customers—a measure of task difficulty—or for worker tenure, as these variables represent mechanisms by which DHH and non-disabled workers may differ.[15] In the Appendix, we show that controlling for restaurant-customer distance has little effect on results. However, not controlling for tenure is substantively important: as we discuss below, worker efficiency and output improve with tenure, and DHH workers are less likely to quit, so higher tenure is a way by which DHH workers benefit themselves and the Platform.

The purpose of equation (2) is to provide descriptive evidence on differences between DHH and non-disabled workers. It is not to estimate the "causal effect" of being DHH.

For each variable besides bad ratings and attrition, we use an inverse-hyperbolic sine transform on y, allowing us to estimate variables in percentage terms. Since we restrict our

[15] These variables can be argued to be "bad controls" if they are mechanisms by which the workers differ (Angrist and Pischke, 2009). Still, it can be interesting to address the question of whether DHH and non-disabled workers differ in efficiency at a given level of tenure or for particular routes.

sample to weeks with positive orders, our outcome variables are never equal to zero, and we avoid the "log of zero"-like problem with inverse hyperbolic sine (Chen and Roth, 2023). For bad ratings and attrition, we estimate linear probability models.

In addition to overall differences, we also estimate outcome-tenure curves separately for DHH and non-disabled workers. This allows us to better understand how outcomes vary with tenure, as well as to understand how DHH differences vary by tenure, e.g., whether differences are persistent or whether there is convergence.

2.2 Results

Figure 1 visually summarizes performance gaps between DHH and non-disabled workers across key outcomes prior to the introduction of AI tool. Each bar reports the percentage difference associated with the DHH indicator estimating equation (2) with full controls. We also discuss results without controls at various times—see Figure A.9 for raw means on several outcomes between DHH and non-disabled.

Productivity: speed variables. Figure 1 shows that DHH workers take longer per delivery than non-disabled workers, though differences vary considerably by stage of the delivery process. For the stage from order initiation by the consumer to order assignment or acceptance by the worker, DHH workers take approximately 2.5% longer than non-disabled workers. For the stage from order acceptance from confirming the order assignment to departure—which includes traveling to the restaurant, communicating with restaurant staff, and picking up the order, there is no difference betwen DHH and non-disabled workers. DHH workers are actually 0.6% faster, but it is highly statistically insignificant. The largest gap occurs in the final stage, from leaving the restaurant to delivering the order to the customer: DHH workers take 4.2% longer, which is about one minute more. Overall, DHH workers do 1.6% fewer orders per service hour than non-disabled workers.[16]

Figure 2 presents average times for each of the three delivery segments as a function of worker tenure, showing simple means over tenure with no control variables.[17] Time spent in every segment of the order fulfillment process declines with tenure for both groups, consistent with learning by doing (Shaw and Lazear, 2008). For the first two segments—time to order acceptance and time from acceptance to store pickup—some differences are present in the first three months of tenure, but the two lines converge starting in the 4th month of tenure.

[16]Orders per service hour is calculated using $\frac{60}{t_1+t_2+t_3}$, where the denominator is the sum of the averages of time in minutes. This calculation doesn't use our hours data, which are based on hours available for work as opposed to hours in transit on an order.

[17]It is also consistent with differential attrition, where less productive workers quit as they learn their productivity, gradually leaving the highest productivity drivers. We discuss differential attrition later.

In contrast, for the final segment going from order pickup at the store to customer delivery, the gap is more persistent. That is, not only is the difference in delivery time largest for time from pickup to customer delivery, but it is also the segment without convergence.

The time numbers on each stage of the delivery process are averages across worker-orders. The Platform may be more concerned with more extreme events, like whether an order is substantially late. Figure 1 shows that DHH workers have 8% more late deliveries, defined as deliveries beyond the expected window, than non-disabled workers. This works out to DHH workers having an extra 33 late deliviers per 10,000 orders.

A natural explanation for DHH workers being slower is communication challenges with customers. Other possibilities include communication challenges with restaurants or differences in general driver behavior (e.g., DHH workers driving slowly or more cautiously). However, only communication challenges with customers explain why DHH drivers are slowest on the final stage of the delivery.

This mechanism dovetails with what we observe in our qualitative fieldwork, especially riding along drivers. On most rides, DHH workers seem to go as fast as non-disabled workers. However, there are occasional situations that require extensive communication with customers, leading to substantial delays. In China, delivery workers are generally expected to call the customer upon arrival, e.g., to coordinate the handoff or confirm delivery if the food is left in a delivery cabinet or at the front desk. Prior to the AI tool studied, DHH workers who couldn't speak had to rely on in-app messaging, which customers sometimes don't see, contributing to occasional but more pronounced delivery delays.

Productivity: quality. As noted above, our measure of work quality is bad customer reviews. Figure 1 shows that DHH workers are 37% more likely to receive a bad review in a week than are non-disabled workers. In terms of the number (i.e., count) of bad reviews, DHH workers receive over 50% more bad reviews than non-disabled workers. These differences are similarly strong for both gig and team workers separately (Figure A.1). Although bad reviews are rare in absolute terms, these are sizable percentage differences between DHH and non-disabled workers in the pre-AI period. They reinforce the above discussion on mechanisms pointing to challenges DHH drivers face in communication with customers.

Our qualitative work supports that communication challenges with customers are a natural explanation for this result as well. Customers may get frustrated if communication is difficult, e.g. if the worker can't place a call to locate the customer, the customer may misinterpret the lack of spoken communication as rudeness. Several DHH workers interviewed recounted situations where customers did not realize the worker was DHH and left bad ratings citing "not calling" or "not speaking clearly."

Overall, the results on speed and bad ratings support the assumption in our framework

of $z_D < z_N$.

The presence of bad ratings is a concern for the Platform. If customers are dissatisfied, they may order less frequently, and the Platform could get a bad reputation. The Platform itself noted this concern, which was one motivation for investing in calling technology.

Labor supply. Figure 1 shows that DHH workers work 6.8% more hours per week than non-disabled workers, conditional on controls, equivalent to an extra 2.7 further hours per week. Additional hours could come from DHH workers working more days per week or more hours per day, and our data indicate both. Specifically, DHH workers work an extra 4.4% days per week than non-disabled workers, equivalent to an extra 0.3 days per week.

Beyond working more hours and days, Figure 1 shows that DHH workers are 32% less likely to quit the Platform than non-disabled workers, conditional on observables, in a linear probability model. To dig further into this overall difference, panel (a) of Figure 8 presents Kaplan-Meier survival curves for disabled and non-DHH workers separately. The raw gaps in survival emerge starkly within the first month and remain even after a year.

These are substantial differences in hours and driver retention. To benchmark magnitudes for retention, research across numerous firms shows that workers hired via employee referral are less likely to quit than workers hired through other means, but differences for lower-skill workers are generally on the order of 10-15% (Hoffman and Stanton, 2025). Thus, the retention difference by DHH status is roughly twice that associated with being referred . Appendix Figure A.1 shows that the tendencies of DHH workers to work more hours, work more days, and quit less are all present when one looks separately at gig and team workers.

Our conceptual framework implies that DHH workers should work more if differences in efficiency are small relative to differences in outside options. We believe this is exactly what is occurring in our setting, where Figure 1 indicates that differences in efficiency are small, at least for efficiency as defined in terms of overall speed. In our qualitative work, it was not uncommon for DHH workers to work 6-7 days a week on the Platform, in light of a scarcity of stable alternative job opportunities. Our findings of DHH workers quitting less broadly aligns with Minor et al. (2018), who find that workers with criminal records are less likely to quit, also seemingly because of worse outside options.

Total output. Combining the two offsetting forces—DHH workers being slightly slower that supplying more hours—we find that the net effect is that DHH workers deliver more orders per week than non-disabled workers. On average, DHH workers do 5.7% more orders than non-disabled workers, equivalent to an extra 8 orders per week.

Panel (a) of Figure 6 shows orders as a function of worker tenure in the pre-AI period. As for the speed variables, performance increases with tenure. However, for total output, DHH and non-disabled workers have very similar orders at given tenure levels. The advantage of

DHH workers arises from them having higher tenure and quitting less. That is, DHH workers operate at higher tenure levels of the learning curve than do non-disabled workers.

The output gap is largest in the team model. Among team workers, DHH workers complete 8.8% more orders per week than non-disabled workers, about 18 additional orders. Among gig drivers, the difference is 1.5% and statistically insignificant. A likely explanation is greater support in the team model, where DHH workers can seek help from station managers and teammates.

Income, wages, and profits. Figure 1 shows that, in line with doing more orders per week, DHH workers earn 5.9% more in weekly gross income relative to non-disabled workers.[18] This increase is driven entirely by greater labor supply rather than higher pay per unit of time: hourly wage rates for DHH workers are lower than those of non-disabled workers, as also seen in Figure 1.

Not accounting for bad rating penalties, disabled workers earn approximately 7.6% lower hourly wages. When penalties from bad ratings are incorporated, the hourly wage gap widens slightly to 8.4%. These estimates are in line with those in Forth and Jones (2025), who estimate a disability hourly pay gap of 6.8% within firm for a representative sample of firms in England and Wales. Overall, that DHH workers have more bad ratings only accounts for about 10% of the 8.4% wage gap. Most of the pay gap reflects DHH workers having similar income to non-disabled workers despite working more hours.

Turning last to Platform profits, DHH workers generate 4.3% more profit per week than non-disabled workers. While DHH workers are slower and earn less per hour, Platform profits are based on total orders, as seen in equation (1). That DHH workers are slower primarily hurts themselves instead of the Platform. The Platform also benefits by DHH workers being substantially less likely to quit every week, meaning the Platform is less likely to pay the onboarding costs.

3 Evaluating the AI Intervention

3.1 Origins of the AI tool

The "Intelligent Outbound Call" (IOC) is a text-to-voice calling feature on the app to help DHH workers communicate better with customers. Many DHH workers face challenges in speaking and being understood. Before the IOC feature was introduced, such DHH workers communicated with customers via text-based messages within the Platform's app. This

[18] Weekly gross income is based on earnings received, and does not include penalties for bad ratings or costs from e-bike rental, maintenance, or charging. We account for bad rating penalties in the hourly wage.

posed 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 DHH workers.

Development of the tool. The tool was developed in-house at the Platform by around 3 software engineers during off time over several months. Both of the two major Chinese food delivery platforms emphasize the use of technology for good as part of corporate social responsibility, and engineers at the Platform are encouraged to spend time helping others. In developing the tool, the Platform sought to help DHH workers and increase profits.

How the tool works and how it uses AI. Under the IOC feature, each call begins with an introductory message clearly stating that the caller is DHH (*wú shēng*) 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 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 IOC provides workers with a set of default communication prompts, as well as the option to create and customize prompts tailored to specific situations. Figure D.1 illustrates both the default prompts offered by the platform and examples of customized prompts commonly used by DHH workers. These prompts typically fall into three categories: (1) difficulties locating the customer, such as being denied entry by security or requiring a buzzer code; (2) notifications that the order has been placed at a specific location, such as a food cabinet; and (3) polite requests for positive customer reviews.

The advantages of the IOC 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 DHH workers can quickly address common issues through pre-set 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.

AI implementation. The Platform made multiple announcements of the AI-based tool to DHH drivers via in-app notification.

3.2 Overall Effects of the AI Tool

Estimation strategy and graphical presentation of results. To assess overall impacts of the AI tool, we estimate a difference-in-differences (DiD) model that compares outcomes for DHH workers before and after the introduction of the AI tool, relative to changes over the same period for non-disabled workers, who do not receive access to the tool:

$$Y_{it} = \alpha + \gamma \cdot DHH_i \times Post_t + \lambda_i + \xi_t + \varepsilon_{it}. \tag{3}$$

where $Post_t$ is a dummy equal to 1 if the current week is on or after the introduction of the AI tool in late April 2022. γ measures the differential impact of the tool on DHH workers relative to non-disabled workers. Worker effects λ_i capture fixed differences across workers while time effects ξ_t capture common shocks. Beyond overall DiD results, we (1) show event study results analyzing how outcomes change month-by-month before and after the AI tool and (2) compare outcome-tenure curves before and after the AI tool.

For analysis of employee attrition, we do not include worker fixed effects—as workers can only quit once—and instead use the same controls as for equation (2).

The identifying assumption in (3) is that the error term is uncorrelated with the key regressor, $DHH_i \times Post_t$, conditional on worker effects and time effects. This could be violated if there are other policies or procedures introduced for DHH workers around the event time window (for the event study analysis) or in the sample period in general (for overall effects). We examine the concern about differential trends using event studies.

Figure 3 summarizes DiD estimates of the AI tool's impact across all outcomes. Each bar is an estimate from equation (3) and represents a percentage change impact on the outcome for DHH workers. We analyze the same outcome as in Figure 1.

Productivity variables: speed. Figure 3 shows that the AI tool led to an improvement in DHH worker speed, with orders per service hour increasing by 1.3%. Improvement occur exclusively in the final stage of order fulfillment, delivery from restaurant to customer, where there was the largest initial disadvantage for DHH workers. Time from pickup to delivery decreases by 3.2% for DHH workers, closing roughly three-quarters of the pre-AI gap of 4.2%. In contrast, effects on time until order acceptance, and time from order acceptance to pickup, are economically negligible and statistically insignificant.

In terms of magnitudes, the reduction in delivery time of 3.2% is equivalent to moving about 3 months down the delivery time tenure curve.

Panele (b) of Figure 4 shows event study estimates of the impact of the AI tool on time from pickup to customer delivery. There are not systematic pre-trends, consistent with the identifying assumption in DiD. In the first month of the AI tool, time from pickup to delivery actually increases, consistent with drivers taking a little time to use the tool effectively. Effects gradually increase in magnitude.

The impact of the AI tool on substantially late deliveries is even larger, with a 5% reduction for DHH workers. The IOC closes nearly two-thirds of the pre-AI gap between DHH and non-disabled workers.

As seen in Appendix Figure A.3, reductions in time from pickup to delivery and in late delivery are pronounced and economically similar for both gig workers and team workers. Overall, the AI tool increases orders per service hour by 2.1% for gig workers and by 1.3% for team workers.

Productivity variables: quality. Figure 3 shows that the AI tool has an especially large effect on bad ratings, reducing the probability of a bad rating by 18%. This closely roughly half of the pre-AI gap of 37%.

As seen in Figure 7—which plots bad ratings over tenure with no controls—prior to the AI tool, there are pronounced gaps between DHH and non-disabled workers that persist for most of a worker's first year of tenure on the Platform. After the AI tool, there is still an initial gap in the first months of tenure, but the gap is much smaller, and the gap dissipates by around the 4th month of tenure.

In terms of magnitudes, the impact of the AI tool is equivalent to moving about 3 months down the tenure curve in bad ratings.

Taken together, these results indicate that IOC achieves its intended purpose: it alleviates communication frictions at the delivery stage, reduces delays and customer dissatisfaction, and modestly improves per-hour productivity for DHH workers, while leaving earlier stages of the order fulfillment process largely unchanged. Qualitative evidence from post-IOC interviews supports this interpretation. Workers reported fewer misunderstandings at delivery, greater confidence in interacting with customers, and fewer incidents requiring extended clarification. Some workers noted that customers reacted positively upon realizing that the call was automated and designed to assist hearing-impaired workers, occasionally leaving supportive feedback.[19]

Labor supply. Pre-AI, DHH workers have greater hours and retention than non-disabled workers, and Figure 3 shows that the AI tool further accentuates these differences. The AI

[19]An alternative interpretation is that customer awareness of a worker's disability may induce greater sympathy in ratings. While we cannot fully rule this out, the concentration of effects in delivery-related metrics suggests that reduced communication frictions play a central role.

tool increases DHH weekly hours online by 8.9%, a more than doubling of the pre-AI gap of 6.8%. The AI tool increases work days per week by 3.3%, which is significant, but most of the increase in hours comes from an increase in hours per day. The AI tool further reduces DHH attrition from the Platform by 9.5%, further expanding the pre-AI gap of 32% by nearly one-third. As seen in the comparison of worker survival curves in Figure 8, the AI tool noticeably increases the difference between the DHH and non-disabled survival curves.

Effects on labor supply and retention are somewhat different from gig and team workers, as seen in Appendix Figure A.3. Gig workers' hours increase substantially by 12.4% while their retention is not significantly affected, i.e., there is a boost to the intensive margin, but not their extensive margin. In contrast, team workers' hours are unchanged, but their retention further improves. Team workers have flexibility over hours, but are already closer to full-time to start with. Thus, team workers may naturally react to the AI tool by adjusting their extensive margin, whereas gig workers who can work part-time may naturally adjust intensively.

These effects are in line with the conceptual framework, which predicts that drivers increase labor supply in response to an improvement in efficiency. While our intentionally simple framework does not account for non-pecuniary work amenities, it is also possible that the AI tool increases labor supply "directly" by making Platform work more pleasant instead of simply by helping DHH workers make more money.

Total output. Figure 3 shows that the AI tool increases DHH weekly orders by 11%. Pre-AI, DHH have workers have 6% more orders than non-disabled workers, so the AI massively extends this initial advantage. Effects are present for team workers at 8%, but they are especially large for gig workers at 16%, and recall that, pre-AI, DHH workers only had higher output than non-disabled workers for team workers. These results suggest that the AI particularly helps DHH workers in the more independent gig mode. Team workers may already have had support or workarounds in place in the pre-AI period and benefited moderately, whereas gig workers, with less support, benefited substantially. As seen in panel (a) of Figure 4, effects are present directly after the AI tool is introduced, and seem to increase over time, likely reflecting in part that effects on delivery speed build over time.

Income, wages, and profits. Following the introduction of IOC, weekly gross income earned by DHH workers increases by approximately 11.3% relative to non-disabled workers. This increase reflects a combination of higher labor supply and improved task-level efficiency, consistent with the earlier results on hours worked, delivery performance, and output.

Hourly wages also rise for DHH workers, though by a more modest magnitude. Excluding penalties, hourly earnings increase by approximately 5.1%, while hourly earnings including penalties increase by about 5.0%. These estimates imply that roughly half of the increase in

weekly income is driven by higher hourly pay, with the remainder attributable to increased hours worked. Given a pre-intervention disabled pay gap of approximately 8%, these gains imply that IOC closes more than half of the wage gap between DHH and non-disabled workers. We are not aware of other workplace interventions or technologies that generate effects of comparable magnitude on disability wage differentials.[20]

From the platform's perspective, the profitability of DHH workers rises substantially following the introduction of IOC. Net weekly profit generated per DHH worker increases by approximately 11.3%, a sizable and statistically significant effect. Because the tool was developed internally by platform engineers, there are no per-use licensing or subscription costs. Even under conservative assumptions regarding development and maintenance costs, the private returns to the platform from IOC appear to far exceed its implementation costs.

In summary, the Intelligent Outbound Calling feature delivers sizable gains along multiple dimensions. It increases efficiency and service quality, expands labor supply, boosts total output, and raises both worker earnings and platform profits. By directly alleviating communication frictions at the point of delivery, IOC substantially narrows pre-existing performance and wage gaps between DHH and non-disabled workers.

3.3 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.10) 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

[20] Evaluating the expansion of work-from-home arrangements during the COVID-19 pandemic, Bloom et al. (2025) show that disabled workers experienced larger wage declines in occupations with greater work-from-home exposure, consistent with increased labor supply exerting downward pressure on wages.

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.

3.4 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 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.4, 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.4, 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.

Figure A.6 shows longitudinal trends in missing-hour data. Missing records appear to be driven by platform-specific record-keeping protocols and are unrelated to the implementation of AI. There is no significant difference between DHH and non-disabled workers in the frequency of missing records, indicating that data attrition reflects centralized institutional practices rather than worker behavior or AI-driven interventions.

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.5, 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.

Would the AI tool have equally benefited non-disabled workers? The AI tool

is only introduced to DHH workers, raising the question of whether non-disabled workers would benefit just as much? The heterogeneity results by disability severity suggest this is highly unlikely to be the case. Less-severely impaired DHH workers benefit much less than severely impaired DHH workers, suggesting that non-DHH workers might also benefit much less or perhaps not at all.[21]

4 Conclusion

We comprehensively examine the relative workplace outcomes of DHH food-delivery workers in China, and analyze how they are affected by a disability-focused AI tool. Before the tool is introduced, DHH workers are more likely to be late and get bad customer ratings, though they also work more hours, deliver more orders, and are substantially less likely to quit, patterns consistent with our conceptual framework where disabled workers have worse outside options. While DHH workers earn great profit for the Platform on a per-worker basis, they earn about 8% less per hour.

The introduction of a targeted non-generative AI tool—the Intelligent Outbound Calling feature—substantially enhances performance. Delivery speeds improve, particularly in the "last-mile" stage involving customer interaction. Orders rise and customer complaints fall. Retention, already higher for DHH workers, increases further. The AI tool decreases the disability wage gap by over half and increases DHH worker profitability. These large benefits occur even though the new tool was relatively simple and inexpensive. Far from being a limitation, the simplicity of the tool illustrate 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.

[21] Of course, we cannot rule out a peculiar non-linearity, where non-disabled workers would also substantially benefit, even though severely impaired DHH workers benefit much more than less severely impaired DHH workers. Our paper is about the impact of a disability-focused AI tool, and is not about the general impact of AI on disabled versus non-disabled workers; thus, the non-linear possibility does not threaten the conclusions of our paper.

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.[22] 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.

[22]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.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton university press, 2009.

Autor, David, "Applying AI to rebuild middle class jobs," Technical Report, National Bureau of Economic Research 2024.

_ **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.

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

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.

Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman, "The Value of Hiring through Employee Referrals," *Quarterly Journal of Economics*, 2015, *130* (2), 805–839.

Chen, Jiafeng and Jonathan Roth, "Logs with Zeros? Some Problems and Solutions*," *Quarterly Journal of Economics*, 12 2023, *139* (2), 891–936.

Chen, M Keith, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen, "The value of flexible work: Evidence from Uber drivers," *Journal of political economy*, 2019, *127* (6), 2735–2794.

Chen, Shi, Xiaodong Wang, Weijun Li, Jingao Zhang, Yuge Qi, Jiaqi Teng, and Zhihan Zeng, "Silent Delivery: Practices and Challenges of Delivering Among Deaf or Hard of Hearing Couriers," in "Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems" CHI '24 Association for Computing Machinery 2024.

Chen, Yanyou, Yao Luo, and Zhe Yuan, "Driving the Drivers: Algorithmic Assignment in Ride-Hailing," *Available at SSRN 4299499*, 2024.

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

Cook, Cody, Rebecca Diamond, Jonathan V. Hall, John A. List, and Paul Oyer, "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers," *Review of Economic Studies*, 2021, *88* (5), 2210–2238.

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.

Gernsbacher, Morton Ann, "Editorial Perspective: The Use of Person-first Language in Scholarly Writing May Accentuate Stigma," *Journal of Child Psychology and Psychiatry*, 2017, *58* (7), 859–861.

Grech, Lisa B., Donna Koller, and Amanda Olley, "Person-first and Identity-first Disability Language: Informing Client Centred Care," *Social Science & Medicine*, 2024, *362*, 117444.

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.

Krieger, Linda Hamilton, *Backlash against the ADA: Reinterpreting disability rights*, University of Michigan Press, 2010.

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.

Lee, Sooyeon, Bjorn Hubert-Wallander, Molly Stevens, and John M Carroll, "Understanding and designing for deaf or hard of hearing drivers on Uber," in "Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems" 2019, pp. 1–12.

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.

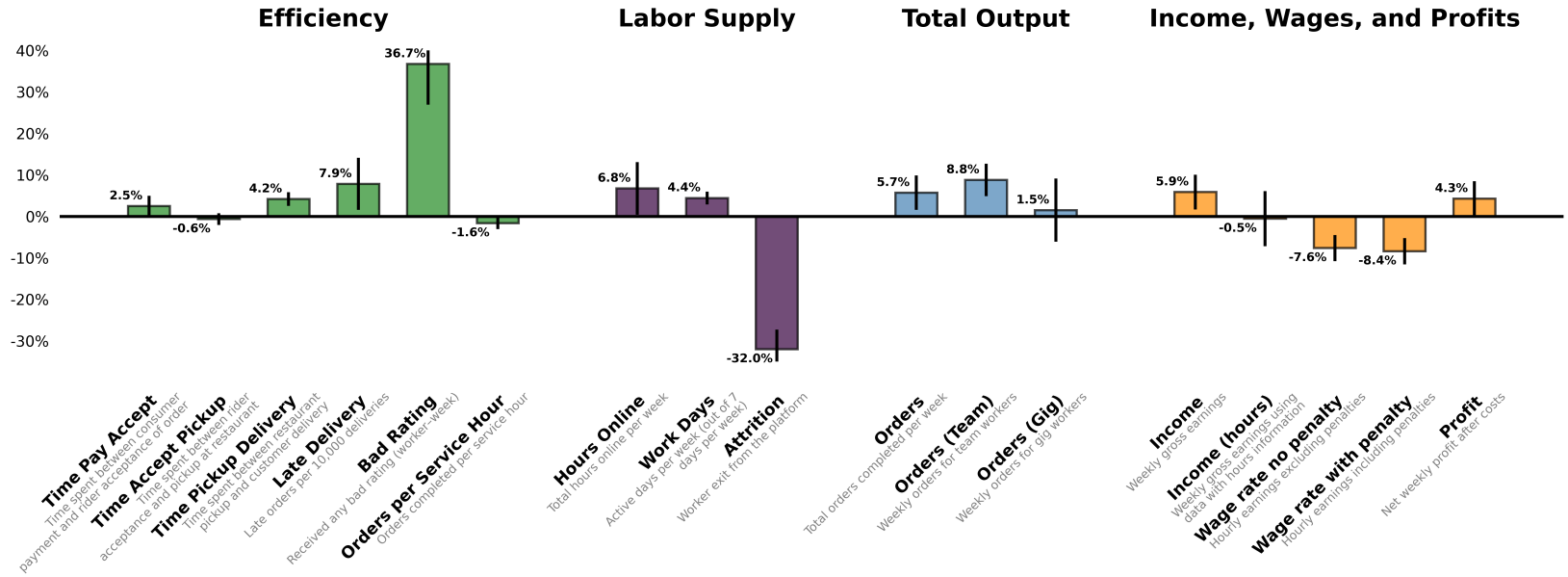
Pallais, Amanda and Emily Glassberg Sands, "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 2016, *124* (6), 1793–1828.

Shaw, Kathryn and Edward P. Lazear, "Tenure and Output," *Labour Economics*, 2008, *15* (4), 704–723.

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.

Figure 1: Pre-AI Normalized Differences in Workplace Outcomes Between DHH and Non-Disabled Workers



Main notes: This figure visually summarizes gaps between DHH and non-disabled workers during the pre-AI period. Each bar represents a percentage difference associated with being a DHH worker for a given outcome. Each bar graph comes from estimating Equation (2) with a different outcome, with the coefficient shown in the bar and the 95% confidence interval accounting for clustering by driver shown in the line.

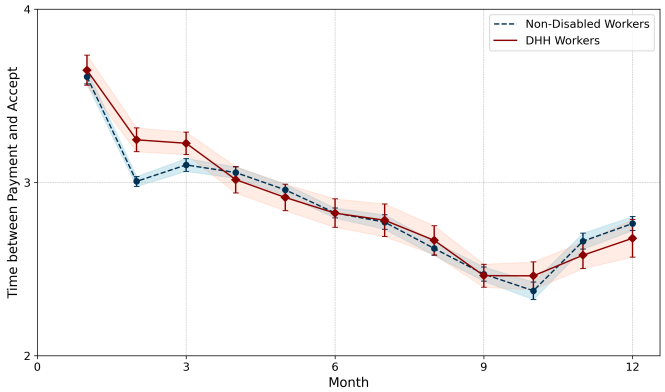
Labor supply notes: In the purple bars, *Hours Online* (total time the worker was online and available for orders, measured in minutes per week), *Work Days* (number of days the worker was active in that week, out of 7), and *Attrition* reflects each worker's likelihood of exiting the Platform.

Efficiency notes: In the green bars, *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. *Late Delivery* represents the number of orders delivered after the scheduled arrival time, normalized per 10,000 orders. *Bad Rating* is an indicator variable equal to one if a worker receives at least one negative customer rating in a given worker–week, and zero otherwise. *Orders per Service Hour* is orders completed per hour online.

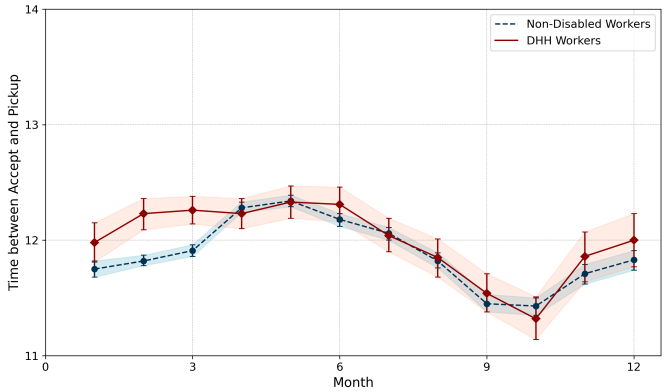
Total output notes: In the blue bars, Orders is the total number of delivery orders completed by a worker in a given week. We also show Orders results separately for gig and team workers.

Wage rate and profit notes: In the orange bars, *Income* is a driver's weekly gross earnings, consisting of regular earnings and subtracting any penalties. Since our hours data are missing for some drivers, we also show results on income restricted to driver's with non-missing hours in *Income (hours)*. *Wage Rate No Penalty)* is our base wage variable and is equal to income divided by hours. *Wage Rate (With Penalty)* is similar to *Wage Rate No Penalty)*, but excludes penalties from income. *Profit* is profit for the Platform.

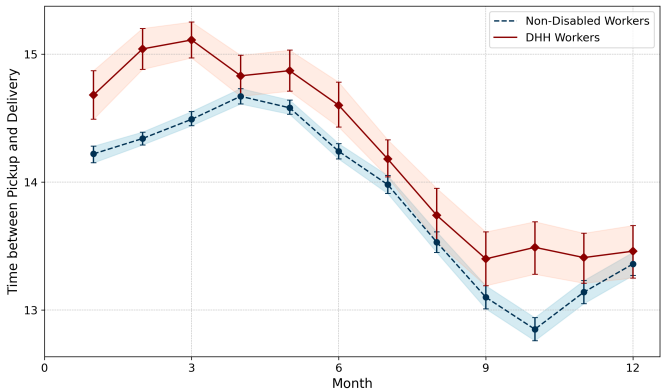
Figure 2: Pre-AI Tenure-Productivity Curves: Time Spent on Different Parts of the Order on Months of Tenure



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



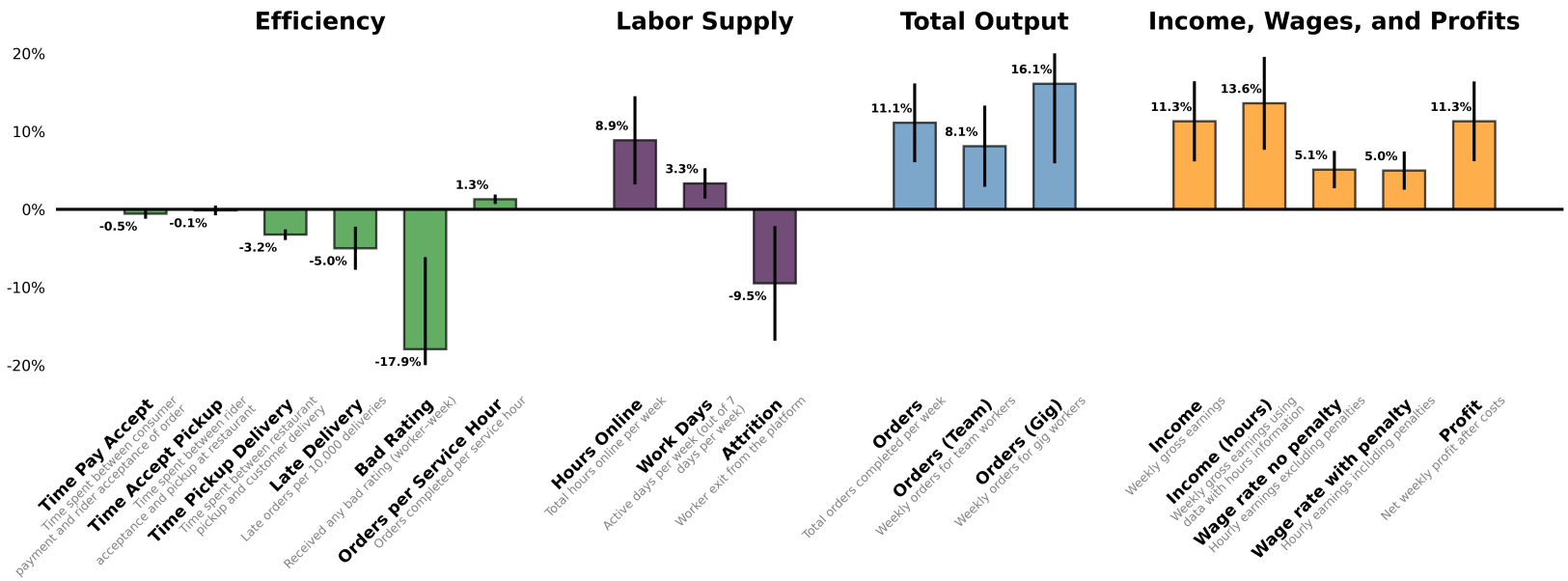
(b) Time Spent Between Worker → Store (Per Order)



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

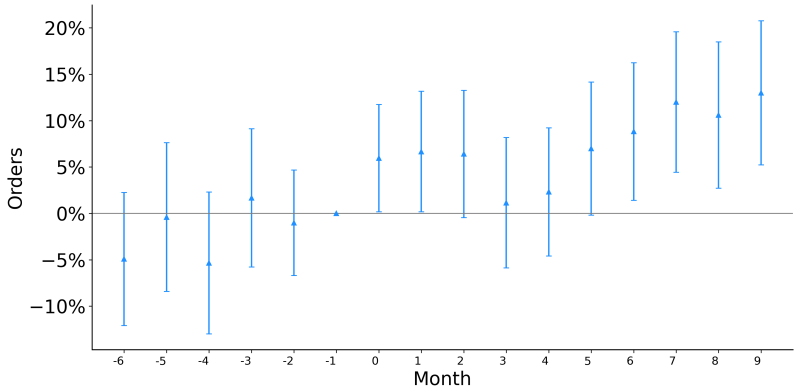
Notes: Figure 2 examines how worker efficiency (speed) changes with experience (*tenure*) on the platform. It plots the average time spent per order at each month of a new worker's tenure for both disabled and non-disabled groups across three panels: **Panel (a) Order → Worker**: Average dispatch time (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 DHH Workers

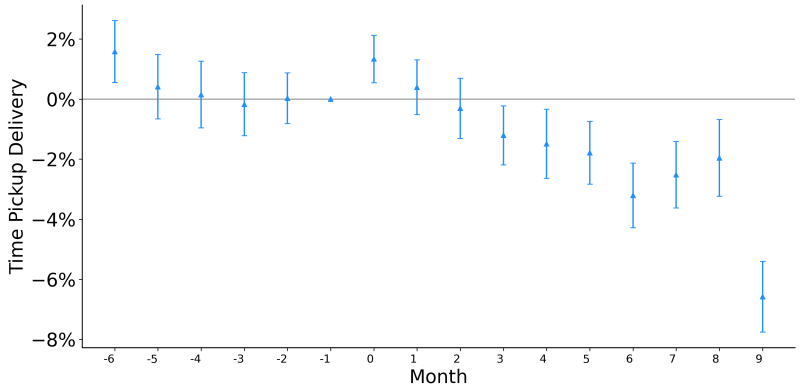


Notes: This figure visualizes the difference-in-differences estimates of the impact of the Intelligent Outbound Call intervention on DHH workers relative to non-disabled workers (Table A.4). Each bar reports the percentage impact of the AI tool on gaps between DHH and non-disabled workers. Positive values indicate improvements for DHH workers relative to non-disabled workers (e.g., higher output or earnings). For duration-based outcomes and negative performance indicators (e.g., delivery times, bad ratings), negative values represent improvements. Error bars denote 95% confidence intervals, with standard errors clustered by workers. Outcome definitions and outcome categories follow Figure 1.

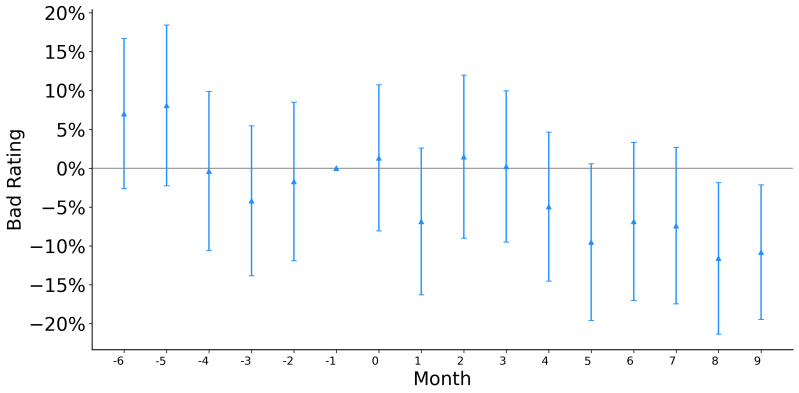
Figure 4: Event Study Analysis: Productivity



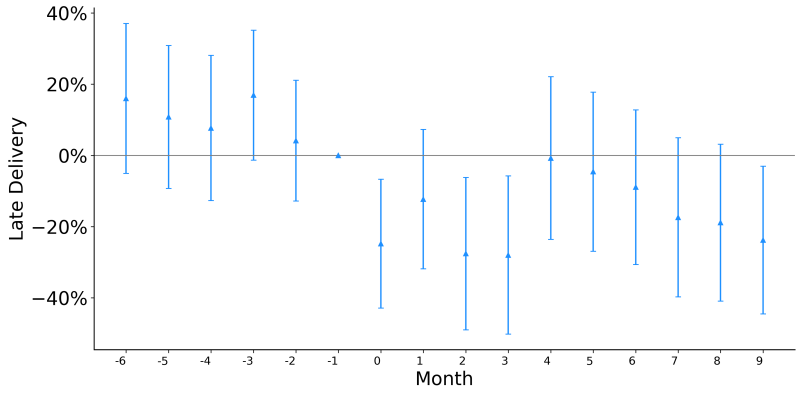
(a) Number of Orders Completed



(b) Time Spent Between Pickup and Delivery (Per Order)



(c) Bad Ratings



(d) Number of Delayed Orders

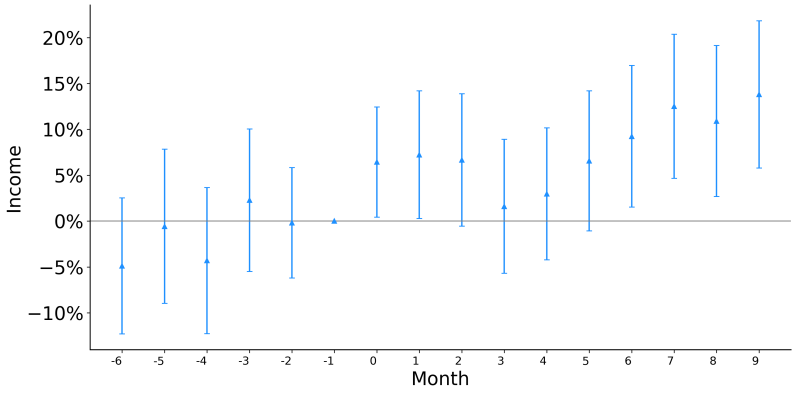
Notes: Figure 4 presents an event study analysis of the productivity impacts of the disability-focused AI tool, namely, the Intelligent Outbound Call (IOC) system. Dots show the event study coefficients, and whiskers show the 95% confidence intervals accounting for clustering by driver.

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 shows the "month" relative to the implementation. Our data are weekly and we define months in terms of 4-week periods. The y-axis shows the estimated difference in orders between DHH and non-disabled workers.

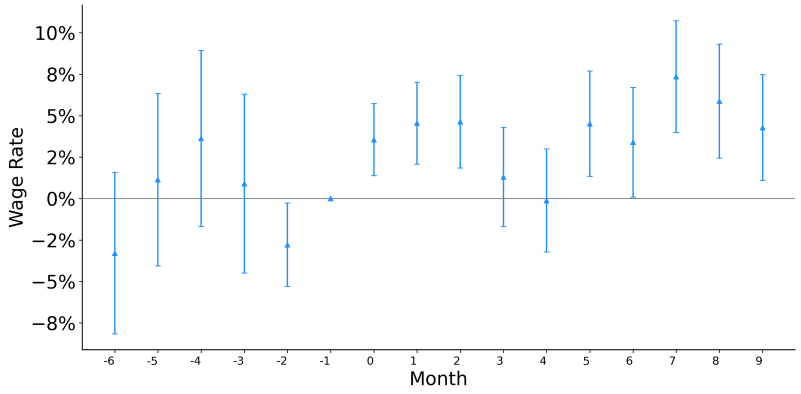
Panel (b) shows the gap in delivery time for the final stage (*Store* → *Customer*) between DHH and non-disabled workers, plotted relative to the implementation of IOC.

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

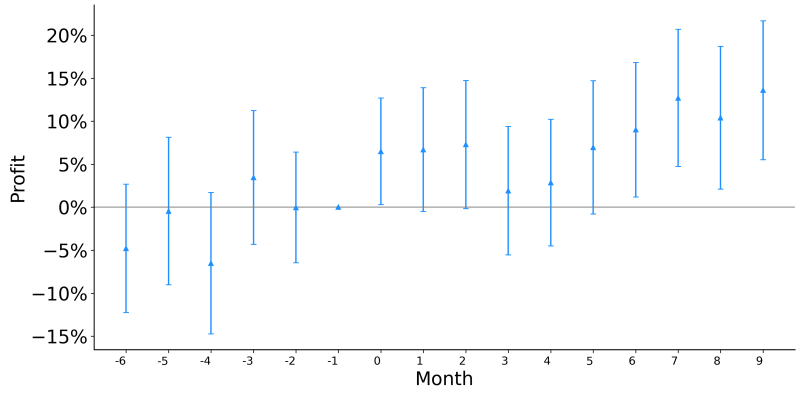
Figure 5: Event Study Analysis: Wage and Profits



(a) Income



(b) Wage



(c) Profit

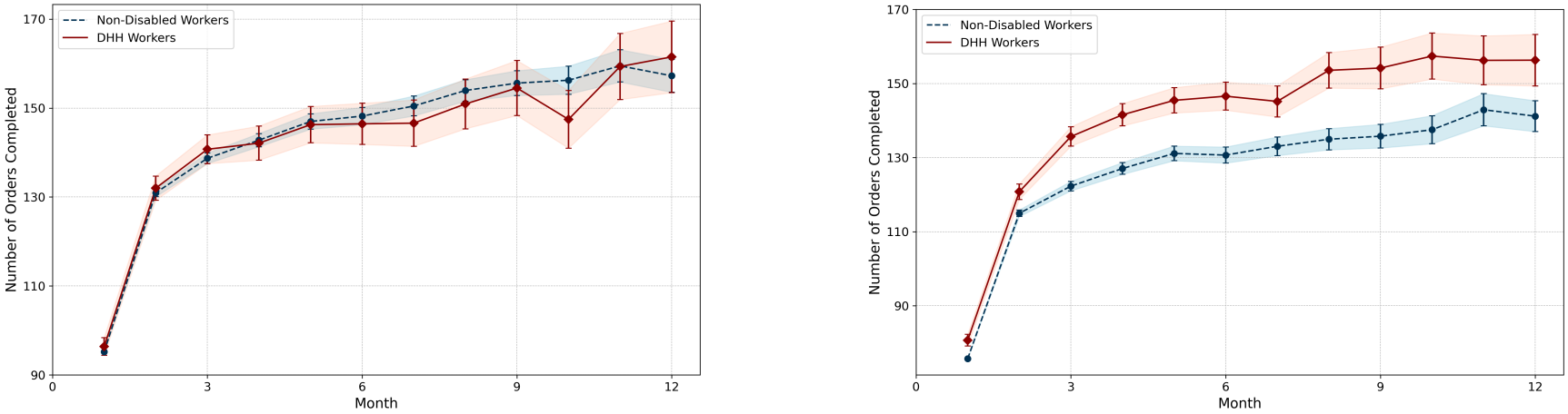
Notes: Figure 5 presents an event study analysis of the introduction of the Intelligent Outbound Call (IOC) system, examining trends before and after the intervention for outcome measures related to wages and profits.

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 the AI Tool



(a) Pre-Periods

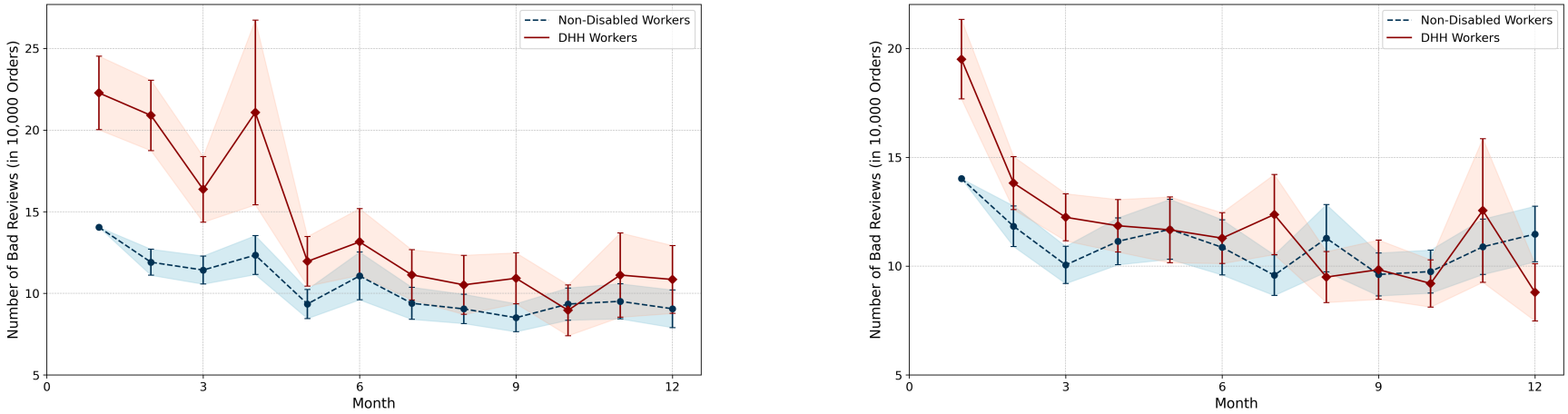
(b) Post-Periods

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. The figure plots the average number of orders completed per month of tenure. Error bars represent the 5th–95th percentile confidence interval. The red solid line corresponds to disabled workers, and the blue dashed line corresponds to non-disabled workers. Tenure profiles are constructed following the same cohort-based approach as Figure 2, but are stratified by whether workers had access to IOC from the start of their tenure. IOC was implemented in April 2022; the pre-period covers January 2021 to early 2022, and the post-period begins in April 2022 and follows workers for up to 12 months.

Panel (a) Pre-Period: Displays tenure profiles for workers who entered the platform prior to the introduction of IOC. For each tenure month (1–12), the panel reports the average monthly order count for disabled and non-disabled workers. The sample includes 191,139 worker-month observations from 17,732 unique drivers, of whom 1,624 are DHH and 16,108 are non-disabled.

Panel (b) Post-Period: Displays corresponding tenure profiles for workers who entered the platform after IOC was introduced. The panel tracks monthly order counts from tenure month 1 through 12. The sample consists of 186,760 worker-month observations from 25,873 unique drivers, including 2,660 DHH and 23,213 non-disabled workers.

Figure 7: Tenure-Productivity Curves in Bad Reviews, Before and After the AI Tool



(a) Pre-AI tool

(b) Post-AI tool

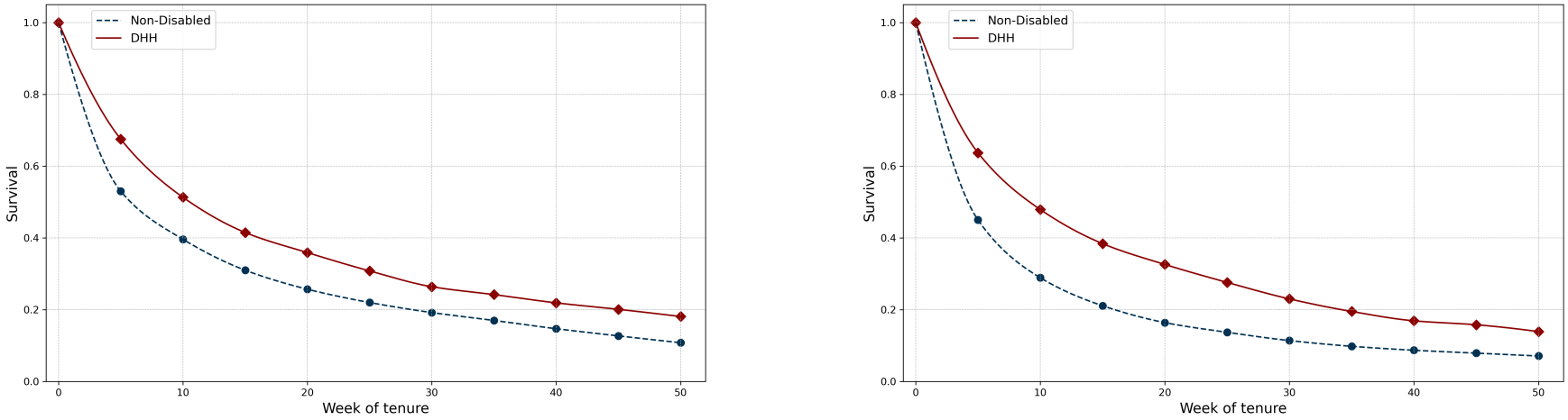
Notes: Figure 7 is analogous to Figure 6, but focuses on service quality, measured by bad customer reviews. The figure illustrates how bad review rates evolve with worker experience and how these tenure profiles differ before and after the introduction of the Intelligent Outbound Call (IOC) technology. The outcome is the number of negative reviews per 10,000 completed orders, computed for each month of tenure. Error bars show 95 percent confidence intervals. The red solid line corresponds to DHH workers, and the blue dashed line corresponds to non-disabled workers. IOC was implemented in April 2022; the pre-period covers January 2021 to early 2022, and the post-period begins in April 2022 and follows workers for up to 12 months.

Panel (a) Pre-Period: Displays tenure profiles of negative review rates for disabled and non-disabled workers who entered the platform prior to the introduction of IOC. For each tenure month (1–12), the panel reports the average number of bad reviews per 10,000 orders.

Panel (b) Post-Period: Displays corresponding tenure profiles for cohorts who entered the platform after IOC was introduced.

The underlying samples and cohort definitions are identical to those used in Figure 6.

Figure 8: Survival Curves: Before and After Comparison



(a) Pre-AI tool

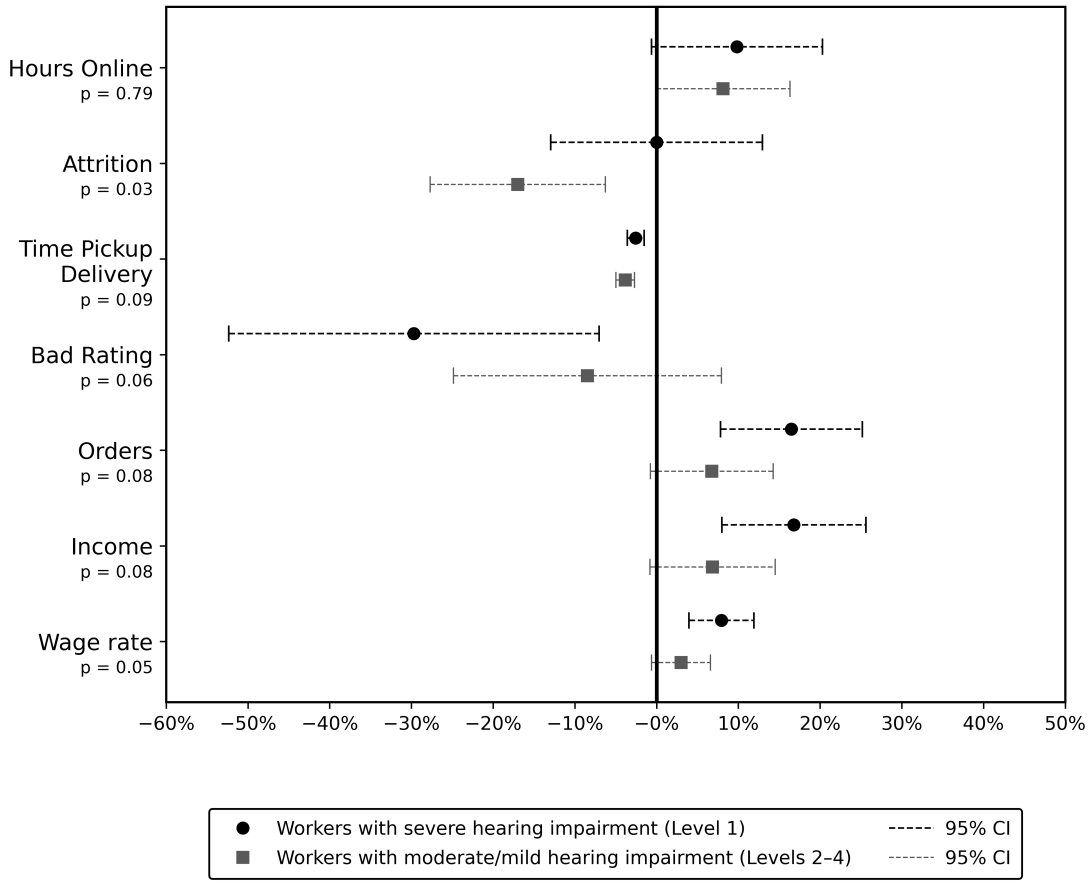
(b) Post-AI tool

Notes: Figure 8 plots survival curves for workers joining the Platform before and after the introduction of the AI tool, doing so separately for DHH and non-disabled workers. The y-axis reports the fraction of workers remaining active on the Platform (i.e., not exited), and the x-axis reports the number of weeks since initial registration. The red solid line corresponds to DHH workers, and the blue dashed line corresponds to non-disabled workers. IOC was implemented in late April 2022. The pre-period covers January 2021 to early 2022, and the post-period begins in April 2022 and follows cohorts for up to 12 months.

Panel (a): Displays survival curves for disabled and non-disabled workers who entered the platform prior to the introduction of the AI tool.

Panel (b): Displays corresponding survival curves for workers who entered the platform after the AI tool was introduced.

Figure 9: Heterogeneous Treatment Effects by Disability Severity



Notes: Figure 9 reports heterogeneous treatment effects by disability severity. Each dot plot is from one regression with DHH workers of a given severity compared to non-disabled workers. The p-values shown are from tests of whether there is a differential effect by disability status for that outcome, and that is from one regression with two key interaction terms. 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.

Table 1

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

	DHH Workers	Non-Disabled Workers	Full Sample	P-Value of DHH vs. Non-Disabled Difference
	Mean	Mean	Mean	
	(1)	(2)	(3)	(4)
Panel A: Demographics				
Age	32.82	30.29	30.54	0.00***
% Male Workers	92.55	90.82	90.99	0.00***
Panel B: Work Model				
% Team Worker	35.36	36.40	36.30	0.18
% Gig Worker	64.64	63.60	63.70	0.18
Panel 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 deaf or hard of hearing (DHH) workers and nearly 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, all at the Worker-Week Level

	Mean (1)	Median (2)	Std. Dev. (3)	Obs. (4)	Workers (5)
Panel A: Efficiency/productivity: Quantity and Quality					
Time from Customer Payment to Order Acceptance, in minutes	3.37	2.48	9.20	464,717	42,281
Time from Order Acceptance to Restaurant Pickup, in minutes	11.97	11.47	6.86	464,850	42,293
Time from Restaurant Pickup to Customer Delivery, in minutes	14.18	13.44	5.10	465,546	42,367
Late Deliveries per 10k orders	416.44	134.23	816.42	480,070	43,605
Bad Ratings per 10k orders	9.62	0.00	94.14	480,070	43,605
Week where driver gets at least one Bad Rating, 0 or 1	0.07	0.00	0.26	480,070	43,605
Panel B: Labor Supply on the Platform					
Hours Online, in hours per week	40.42	37.85	29.19	343,733	34,965
Work Days per week	5.31	6.00	2.13	480,070	43,605
Driver attrites from the Platform, 0 or 1	0.08	0.00	0.27	480,070	43,605
Panel C: Total Output					
Total Orders	152.39	127.00	141.52	480,070	43,605
Peak Hour Orders	90.92	76.00	88.42	480,070	43,605
Total Orders, restrict to Team drivers only	212.35	205.00	146.18	272,986	15,828
Total Orders, restrict to Gig drivers only	73.35	42.00	85.28	207,084	27,777
Panel D: Income, Wages, and Profits					
Driver Income, in CNY per week	Confid.	Confid.	Confid.	480,070	43,605
Driver Income, restrict to weeks with hours data, in CNY per week	Confid.	Confid.	Confid.	343,733	34,965
Hourly Wages, excluding bad rating penalties, in CNY per hour	Confid.	Confid.	Confid.	342,061	34,909
Hourly Wages, inclusive of any bad rating penalties, in CNY per hour	Confid.	Confid.	Confid.	342,061	34,909
Platform Profits from a driver, in CNY per week	Confid.	Confid.	Confid.	480,070	43,605

Notes: This table summarizes weekly work outcomes for workers on the Platform. The unit of observation is a worker-week. Outcome definitions and grouping into the four categories are the same as in Figure 1. Appendix Table A.1 presents summary statistics separately for gig workers and for team workers. The "Confid." marking indicates that certain summary statistics cannot be shared due to confidentiality requirements.

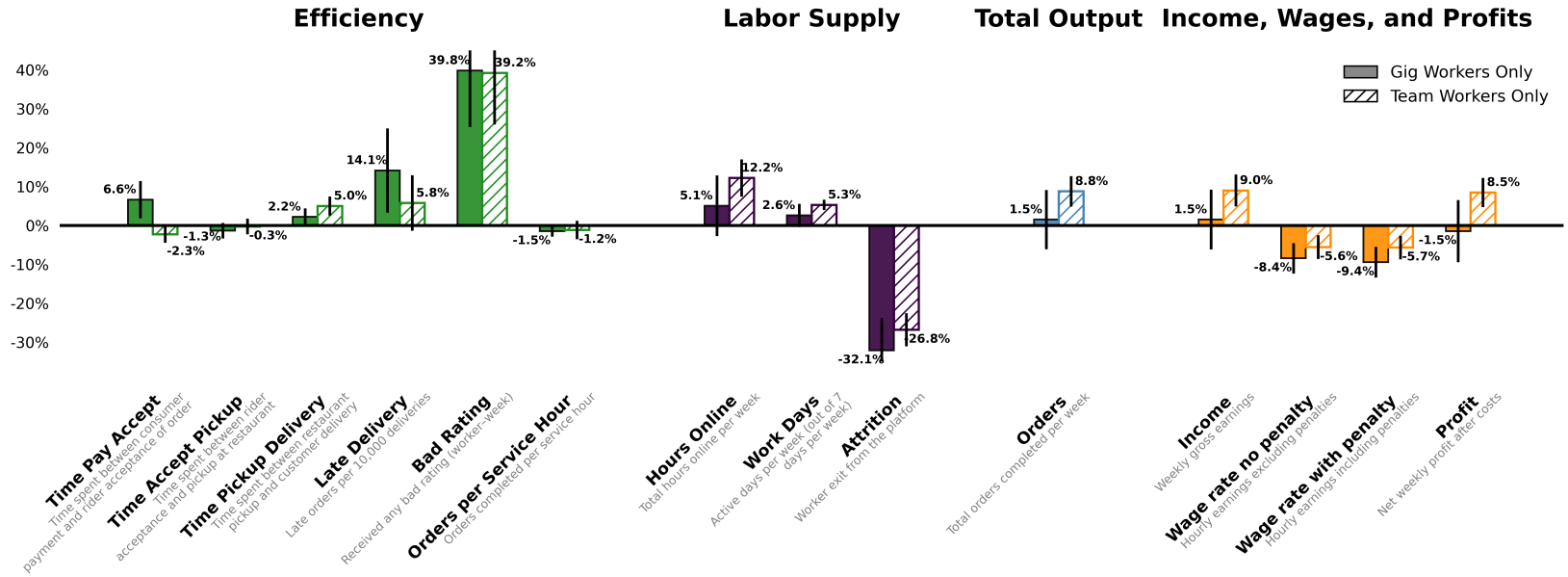
Web Appendix for "Empowering Inclusive Work"

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

Appendix A provides additional figures and tables. Appendix B analyzes how the AI tool affects new signups (i.e., entry) of DHH workers onto the Platform. Appendix C describe the qualitative data collected on DHH workers on the Platform. Appendix D provides further details and visuals on the AI tool.

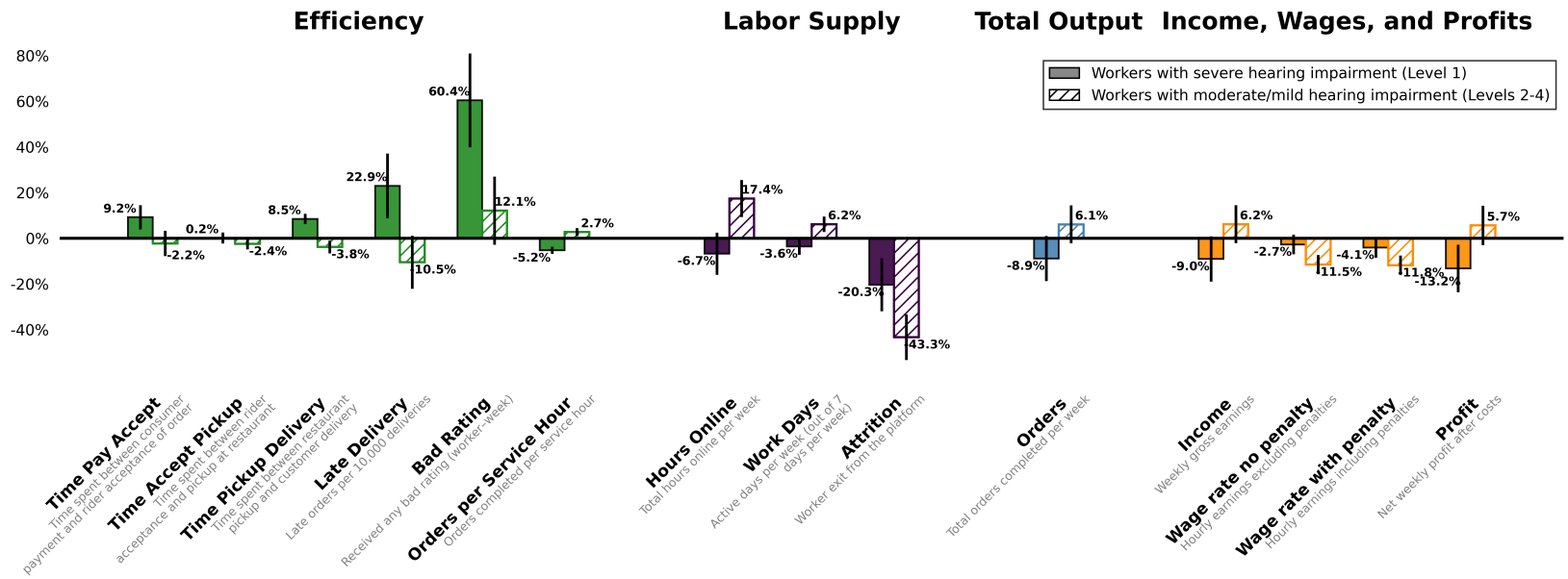
A Appendix Figures and Tables

Figure A.1: Pre-AI Differences Between DHH and Non-Disabled Workers: Gig and Team Workers Separately



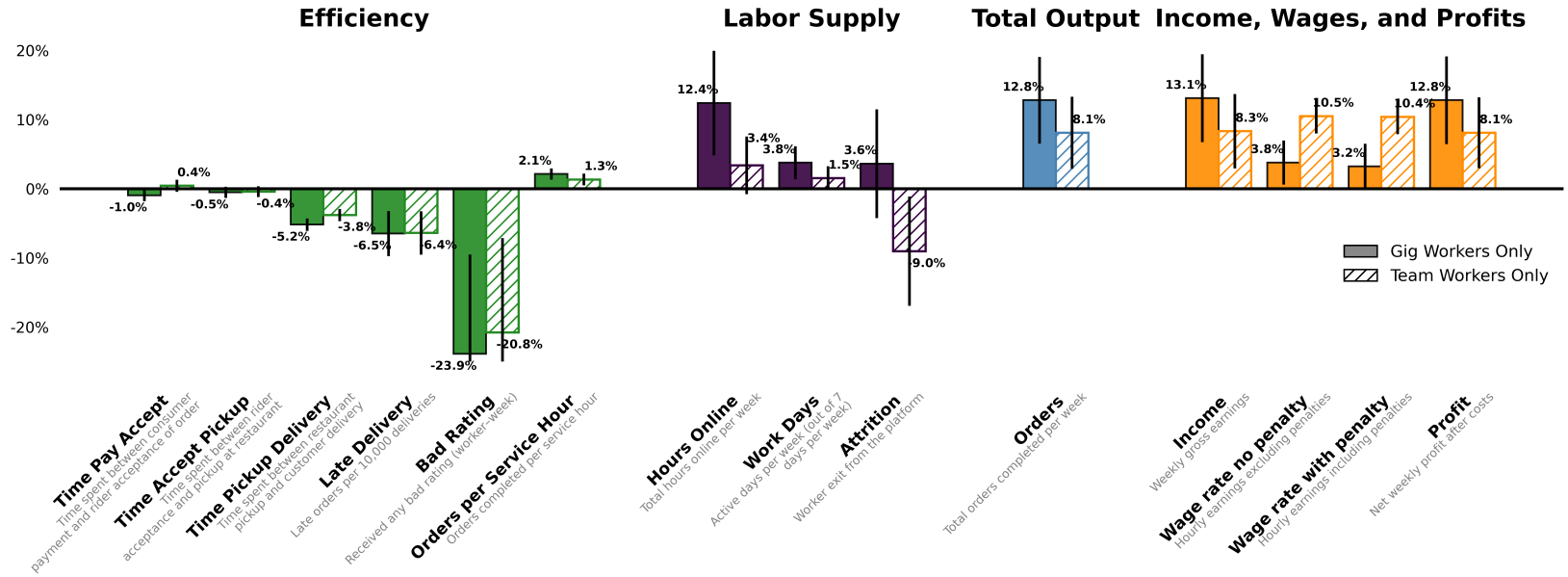
Notes: This figure reproduces the analysis in Figure 1, but splits the sample by gig workers and team workers. We plot differences between DHH gig workers and non-disabled gig workers, as well as differences between DHH team workers and non-disabled team workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept.

Figure A.2: Pre-AI Differences by Disability Severity



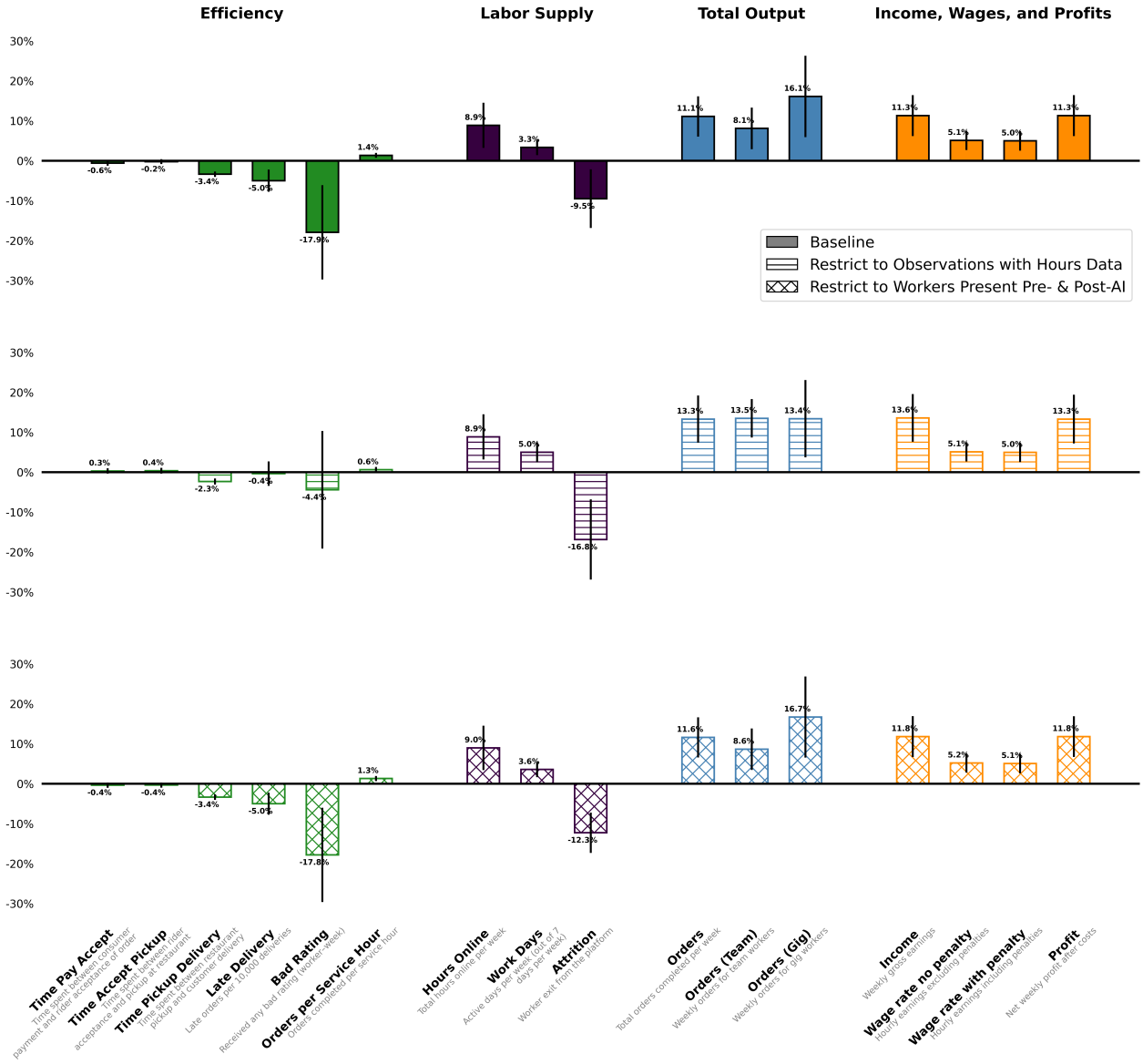
Notes: This figure reproduces the analysis in Figure 1 by disability severity.

Figure A.3: Impact of the AI Tool on DHH Worker Outcomes: Gig and Team Workers Separately



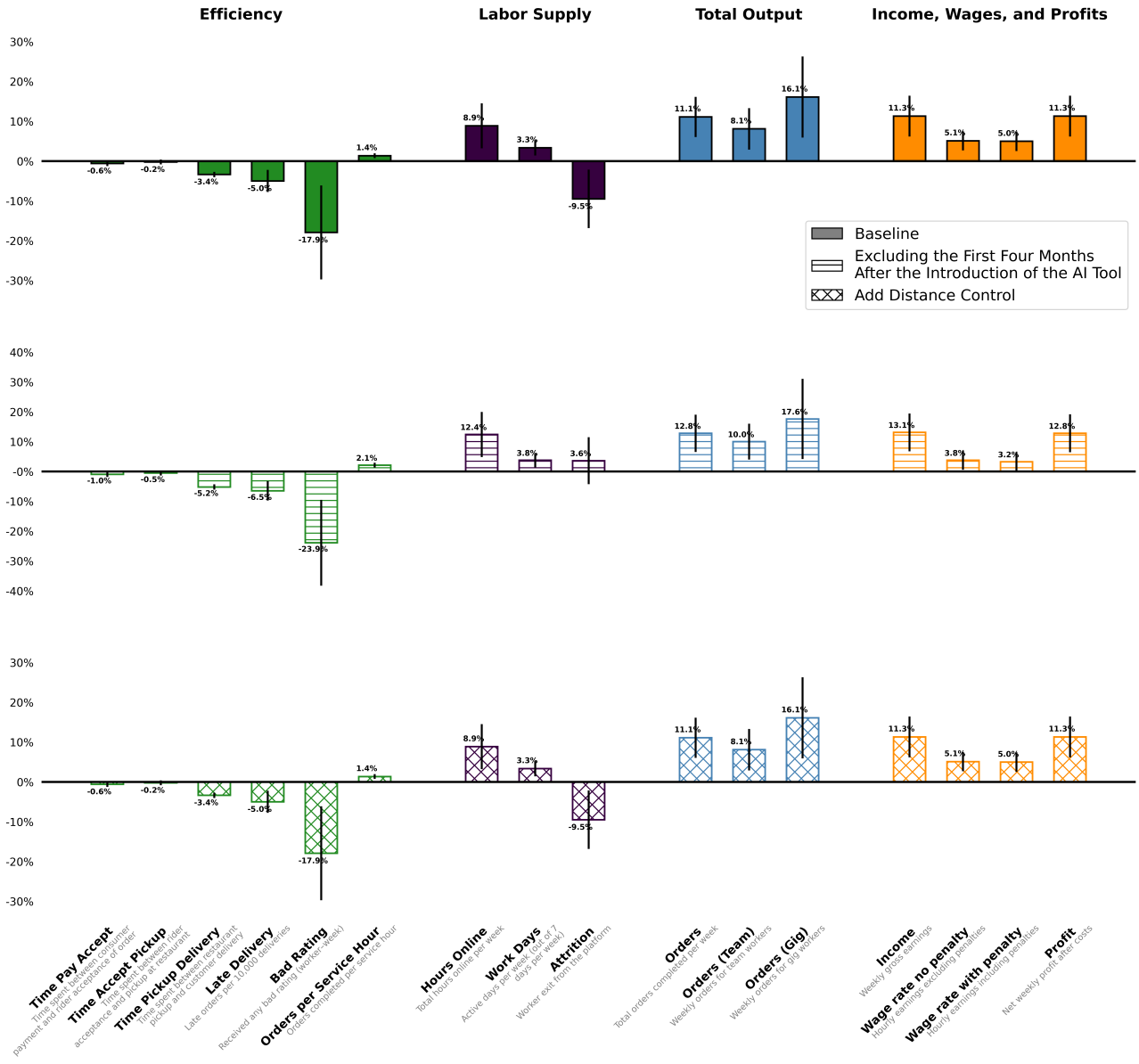
Notes: This figure reproduces the analysis in Figure 3, focusing exclusively on gig or team workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept.

Figure A.4: Robustness of Figure 3 to Sample Selection and Attrition



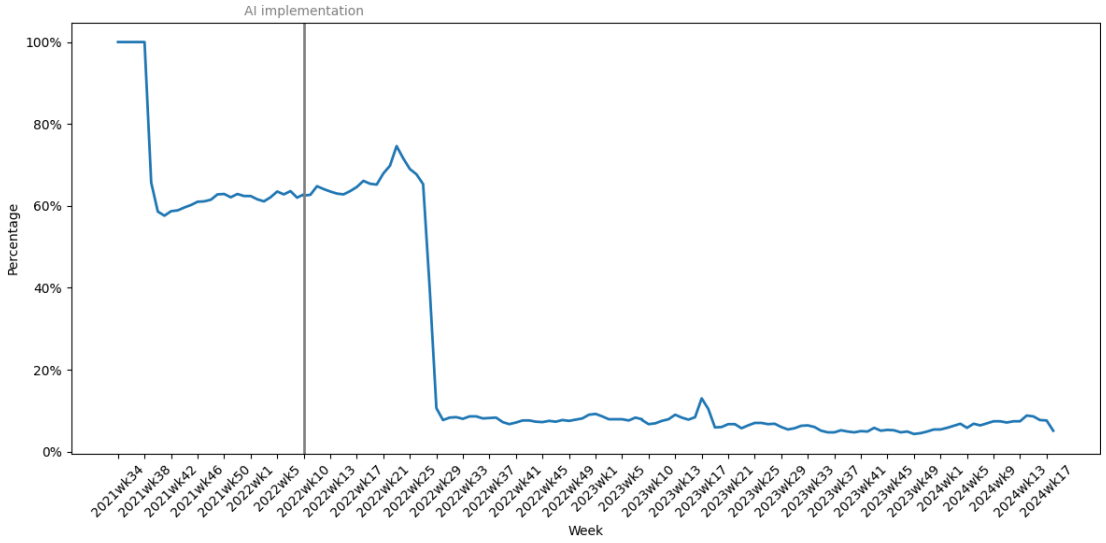
Notes: This figure replicates the analysis in Figure 3. The top panel reports the baseline estimates using the full sample. The middle panel restricts the sample to observations with hours data, addressing concerns related to measurement of labor supply. The bottom panel restricts the sample to drivers present both pre- and post-AI adoption, addressing potential selection and attrition concerns. All estimates are reported as percentage changes relative to the pre-AI mean; bars indicate point estimates and vertical lines denote 95% confidence intervals.

Figure A.5: Robustness of Figure 3 to Experience and Task Difficulty

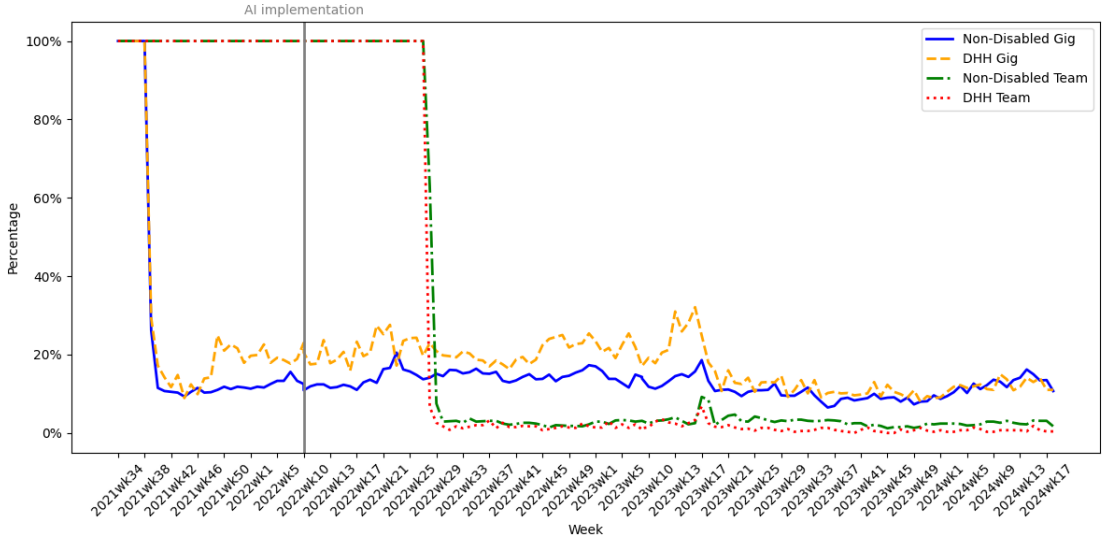


Notes: The top panel reports the baseline estimates. The middle panel excludes the first four months after the introduction of the AI tool for potential learning and onboarding effects. The bottom panel adds controls for task difficulty, proxied by order distance. All specifications follow the same difference-in-differences design as in Figure 3.

Figure A.6: The Share of Worker-Weeks with Missing Hours Data Over Time



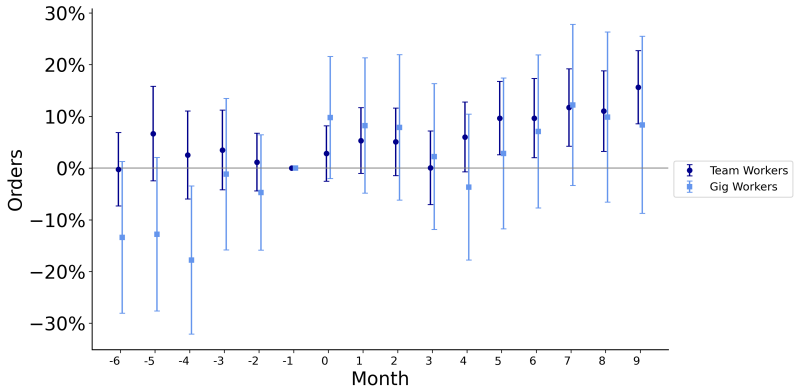
(a) Aggregate percentage of missing hours over time



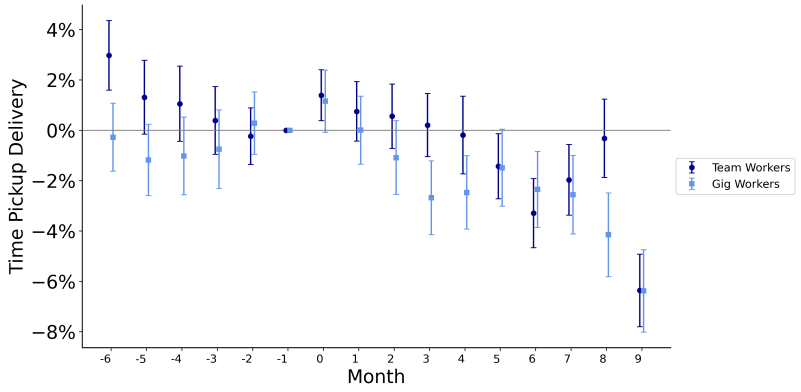
(b) Percentage of missing data categorized by worker type and disability status

Notes: Figure A.6 shows missing-hour patterns by employment type and disability status. The grey vertical marker denotes the timing of AI implementation. Panel (a) shows the overall share of missing hours data over time. As seen in panel (b), which shows missing rate broken down by work model (gig or team) and DHH status, hours data became available for gig workers in mid 2021, and then for team workers in mid 2022.

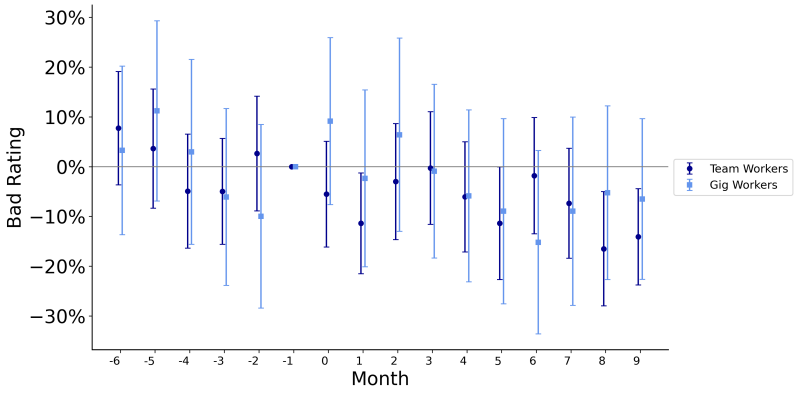
Figure A.7: Event Study Analysis on Efficiency: Splitting by Gig vs. Team Workers



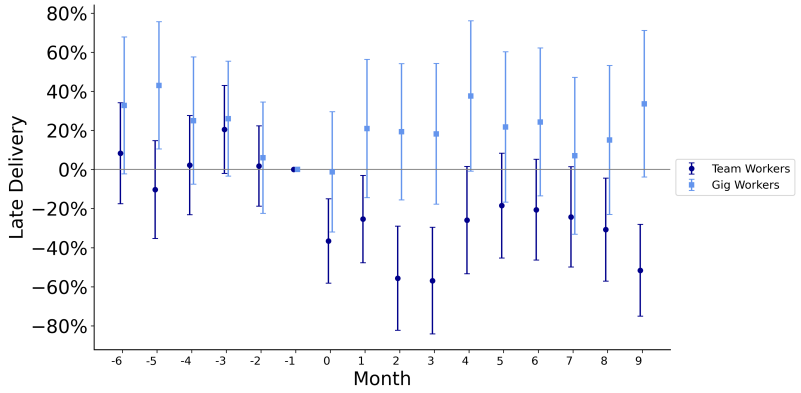
(a) Number of Orders Completed



(b) Time Spent Between Pickup and Delivery (Per Order)



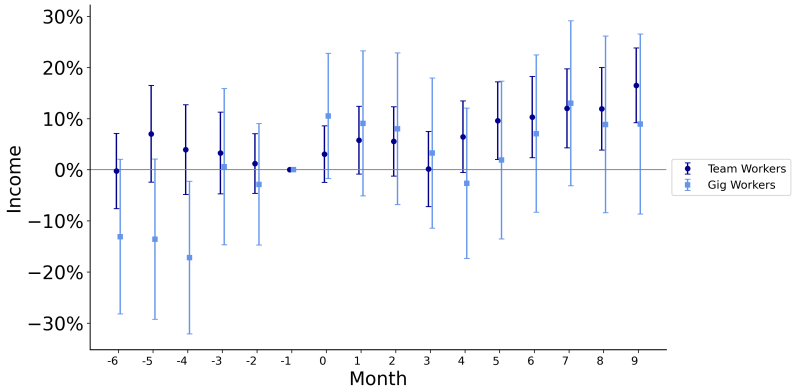
(c) Bad Ratings



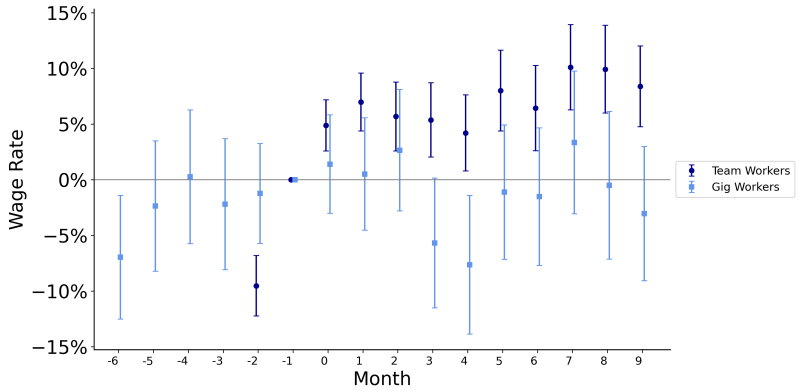
(d) Number of Delayed Orders

Notes: This figure is similar to Figure 6 in the main text, but performs the analysis separately for team workers and for gig workers. The dots show event study estimates, and the whiskers show 95% confidence intervals accounting for clustering by worker.

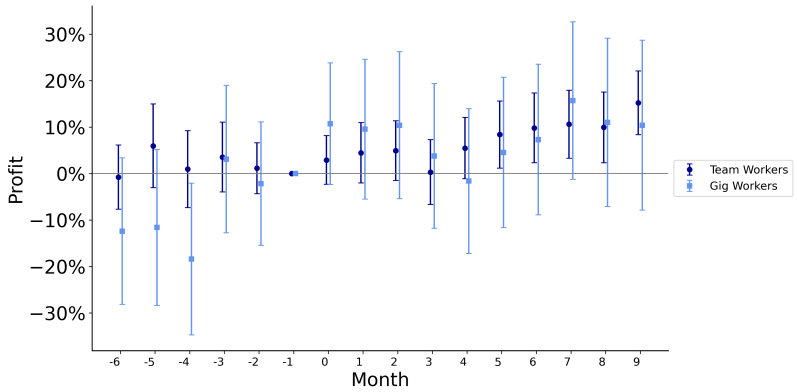
Figure A.8: Event Study Analysis on Wage and Profits: Splitting by Gig vs. Team



(a) Income



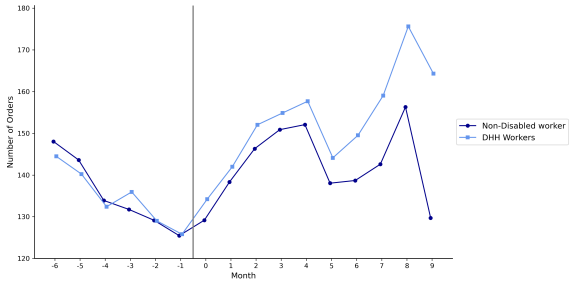
(b) Wage



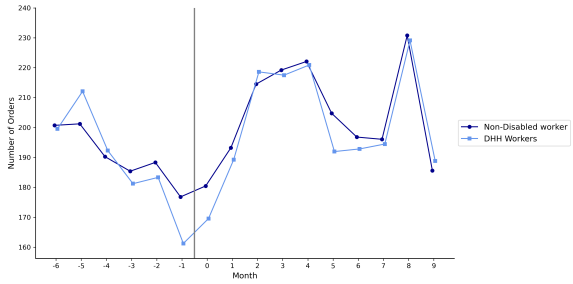
(c) Profit

Notes: This figure is similar to Figure 5 in the main text, but performs the analysis separately for team workers and for gig workers. The dots show event study estimates, and the whiskers show 95% confidence intervals accounting for clustering by worker.

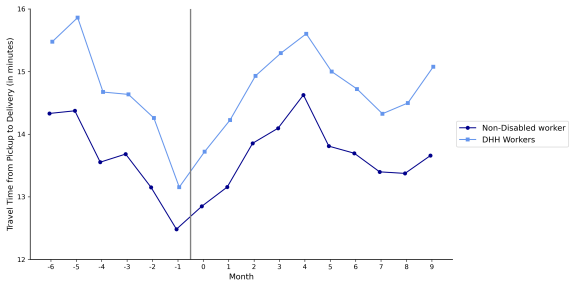
Figure A.9: Comparison of Raw Means Between DHH and Non-Disabled Workers



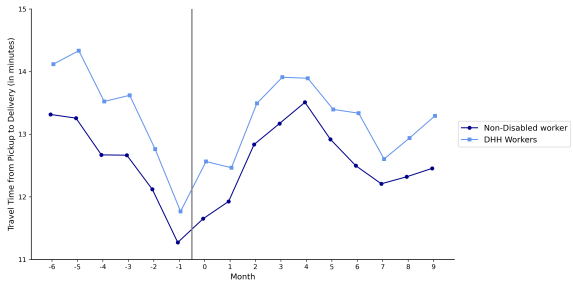
(a) Orders Completed (All Workers)



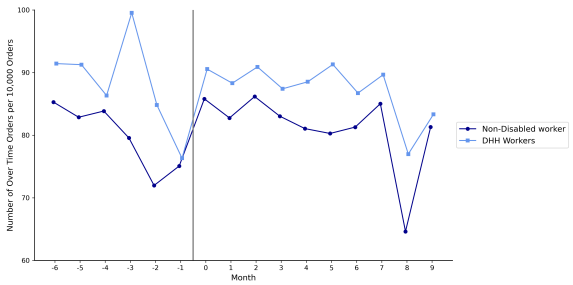
(b) Orders Completed (Non-Exiting Workers)



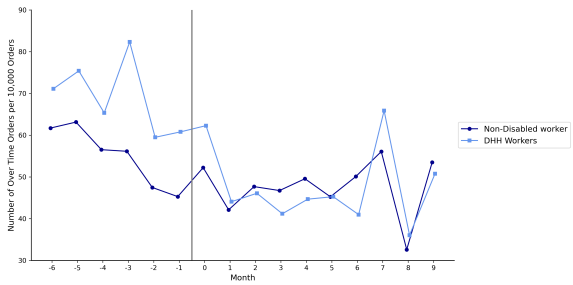
(c) Time Pickup Delivery (All Workers)



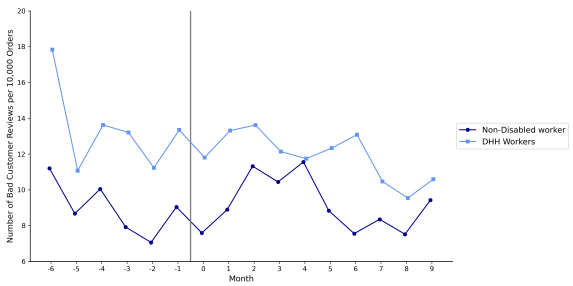
(d) Time Pickup Delivery (Non-Exiting Workers)



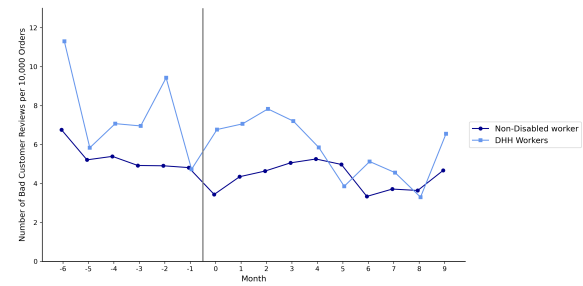
(e) Late Delivery (All Workers)



(f) Late Delivery (Non-Exiting Workers)



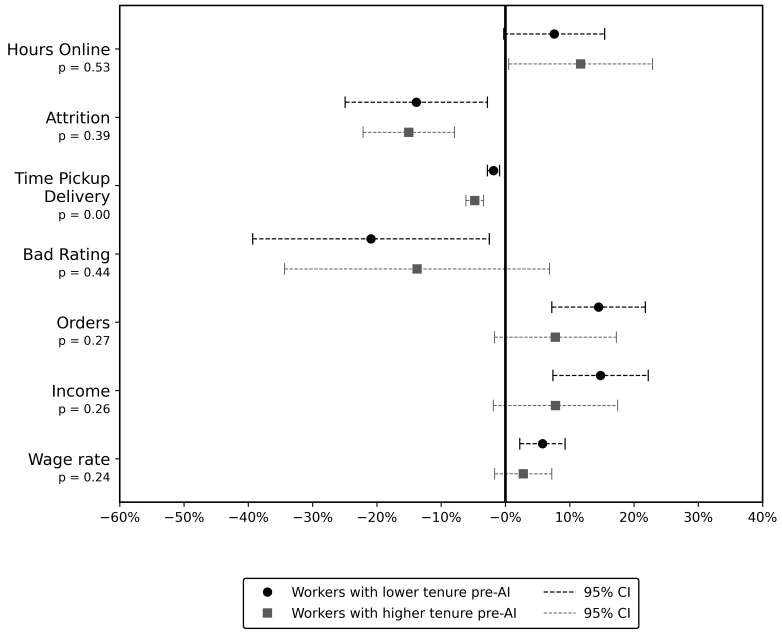
(g) Bad Rating (All Workers)



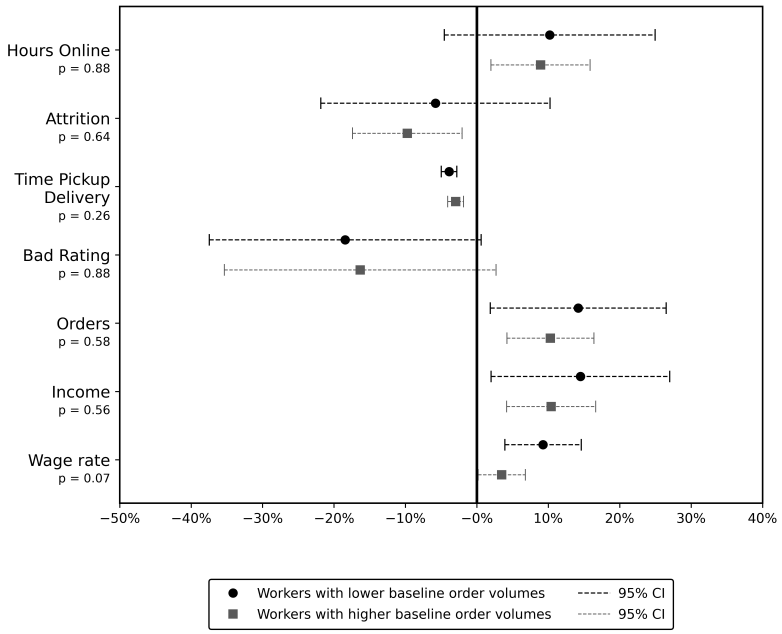
(h) Bad Rating (Non-Exiting Workers)

Notes: Here we compare raw averages of the outcome variables between DHH workers (light blue) and non-disabled workers (dark blue). The left panels show results for all workers. There are 43,065 workers in total, including 4,284 DHH workers and 39,321 non-disabled workers. The right panels show results for workers who remained continuously active on the platform (i.e., did not exit) from month −6 before the AI implementation to month 9 after the AI implementation. This sample includes 11,408 workers in total, with 1,584 DHH workers and 9,824 non-disabled workers.

Figure A.10: Heterogeneity by Pre-AI Tenure and Pre-AI Productivity



(a) Heterogeneity by Pre-AI Tenure



(b) Heterogeneity by Pre-AI Productivity

Notes: Figure A.10 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.

Table A.1: Summary Statistics for Key Outcomes at Worker-week Level: Gig vs. Team Workers

	Gig Workers				Team Workers			
	Mean (1)	Median (2)	SD (3)	Obs. (4)	Mean (5)	Median (6)	SD (7)	Obs. (8)
Panel A: Efficiency/productivity: Quantity and Quality								
Order Acceptance Time, in minutes	4.17	3.29	3.65	206,131	2.74	2.01	11.85	258,586
Pickup Time, in minutes	11.06	10.65	3.55	206,201	12.69	12.07	8.57	258,649
Delivery Time, in minutes	16.69	16.24	5.45	206,237	12.19	11.84	3.76	259,309
Late Deliveries per 10k orders	517.33	140.85	981.70	207,084	339.91	132.16	653.86	272,986
Bad Ratings per 10k orders	15.60	0.00	135.09	207,084	5.09	0.00	41.15	272,986
Week with a Bad Rating, 0 or 1	0.07	0.00	0.26	207,084	0.07	0.00	0.26	272,986
Panel B: Labor Supply on the Platform								
Hours Online, in hours per week	22.83	15.55	21.54	172,695	58.17	61.83	24.83	171,038
Work Days per week	4.25	4.00	2.23	207,084	6.11	7.00	1.65	272,986
Driver attrites from Platform, 0 or 1	0.12	0.00	0.32	207,084	0.05	0.00	0.21	272,986
Panel C: Total Output								
Total Orders	73.35	42.00	85.28	207,084	212.35	205.00	146.18	272,986
Peak Hour Orders	42.87	22.00	54.47	207,084	127.38	125.00	91.75	272,986
Panel D: Income, Wages, and Profits								
Driver Income, in CNY per week	Confid.	Confid.	Confid.	207,084	Confid.	Confid.	Confid.	272,986
Hourly Wages	Confid.	Confid.	Confid.	171,184	Confid.	Confid.	Confid.	170,877
Platform Profits from a driver	Confid.	Confid.	Confid.	207,084	Confid.	Confid.	Confid.	272,986
Unique Workers			27,777				15,828	
DHH Workers			2,769				1,515	
Non-disabled Workers			25,008				14,313	

Notes: This table is similar to Table 2 in the main text, but provides summary statistics separately for gig workers and for team workers. Gig worker outcomes are summarized in columns 1-4, whereas team worker outcomes are summarized in columns 5-8.

Table A.2: Summary Statistics for Key Outcomes at Worker-week Level: DHH vs. Non-disabled Workers

	DHH Workers				Non-disabled Workers			
	Mean (1)	Median (2)	SD (3)	Obs. (4)	Mean (5)	Median (6)	SD (7)	Obs. (8)
Panel A: Efficiency/productivity: Quantity and Quality								
Order Acceptance Time, in minutes	3.96	2.98	10.34	69,444	3.27	2.38	8.97	395,273
Pickup Time, in minutes	11.47	10.94	4.13	69,462	12.05	11.55	7.23	395,388
Delivery Time, in minutes	14.42	13.85	4.93	69,498	14.14	13.36	5.13	396,048
Late Deliveries per 10k orders	426.81	147.06	811.01	70,301	414.67	131.93	817.34	409,769
Bad Ratings per 10k orders	11.53	0.00	83.23	70,301	9.30	0.00	95.88	409,769
Week with a Bad Rating, 0 or 1	0.09	0.00	0.29	70,301	0.07	0.00	0.26	409,769
Panel B: Labor Supply on the Platform								
Hours Online, in hours per week	45.02	46.07	29.05	54,984	39.54	36.16	29.13	288,749
Work Days per week	5.57	7.00	2.18	70,301	5.26	6.00	2.12	409,769
Driver attrites from Platform, 0 or 1	0.05	0.00	0.22	70,301	0.08	0.00	0.27	409,769
Panel C: Total Output								
Total Orders	161.10	141.00	126.66	70,301	150.90	124.00	143.86	409,769
Peak Hour Orders	96.19	85.00	77.27	70,301	90.02	74.00	90.16	409,769
Panel D: Income, Wages, and Profits								
Driver Income, in CNY per week	Confid.	Confid.	Confid.	70,301	Confid.	Confid.	Confid.	409,769
Hourly Wages	Confid.	Confid.	Confid.	54,675	Confid.	Confid.	Confid.	287,386
Platform Profits from a driver	Confid.	Confid.	Confid.	70,301	Confid.	Confid.	Confid.	409,769
Unique Workers	4,284				39,321			

Notes: This table is similar to Table 2 in the main text, but provides summary statistics separately for DHH workers and for non-disabled workers.

Table A.3

Pre-AI Differences Between Disabled and Non-Disabled Workers

	(1) Baseline	(2) Only Gig	(3) Only Team
A. Total Output			
Orders	0.057**	0.015	0.088***
	(0.025)	(0.046)	(0.024)
Orders per Svc Hr	-0.016*	-0.015*	-0.012
	(0.009)	(0.008)	(0.014)
B. Efficiency			
Time Pay Accept	0.025*	0.066**	-0.023*
	(0.015)	(0.029)	(0.013)
Time Accept Pickup	-0.006	-0.013	-0.003
	(0.009)	(0.012)	(0.012)
Time Pickup Delivery	0.042***	0.022*	0.050***
	(0.010)	(0.013)	(0.015)
Late Delivery	31.890**	57.300**	23.420
	(15.450)	(26.750)	(17.600)
Bad Rating	5.115***	9.644***	1.832***
	(1.014)	(2.283)	(0.517)
C. Labor Supply			
Hours Online	0.068*	0.051	0.122***
	(0.039)	(0.047)	(0.029)
Work Days	0.044***	0.026	0.053***
	(0.009)	(0.018)	(0.008)
Attrition	-0.025***	-0.025***	-0.021***
	(0.002)	(0.004)	(0.002)
D. Wage Rate and Profit			
Income	0.059**	0.015	0.090***
	(0.026)	(0.047)	(0.025)
Wage rate (no pen.)	-0.076***	-0.084***	-0.056***
	(0.019)	(0.024)	(0.019)
Wage rate (w/ pen.)	-0.084***	-0.094***	-0.057***
	(0.019)	(0.024)	(0.019)
Profit	0.043*	-0.015	0.085***
	(0.026)	(0.049)	(0.023)

Notes: Standard errors in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Column definitions: (1) Baseline specification using full panel. (2) Sample restricted to Gig workers only. (3) Sample restricted to Team workers only.

Table A.4

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

	(1) Baseline	(2) Only Gig	(3) Only Team	(4) Excl. 1st 4m	(5) Pre/Post	(6) Obs. Hr.	(7) Add. Ctrls
A. Total Output							
1. Orders	0.111***	0.161***	0.081**	0.128***	0.116***	0.133***	0.111***
	(0.031)	(0.062)	(0.032)	(0.038)	(0.031)	(0.040)	(0.031)
2. Orders per Svc Hr	0.013***	0.007	0.012***	0.020***	0.012***	0.005	0.014***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
B. Efficiency							
3. Time Pay Accept	-0.006	-0.007	0.005	-0.009	-0.003	0.004	-0.006
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
4. Time Accept Pickup	-0.001	0.004	-0.003	-0.004	-0.002	0.005	-0.002
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
5. Time Pickup Delivery	-0.033***	-0.022***	-0.037***	-0.050***	-0.032***	-0.022***	-0.034***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)
6. Late Delivery	-20.260***	-5.952	-25.950***	-26.310***	-20.320***	-1.448	-20.260***
	(6.835)	(12.540)	(7.752)	(8.031)	(6.833)	(7.568)	(6.835)
7. Bad Rating	-2.162**	-2.760	-1.642**	-2.308*	-2.122**	-0.998	-2.162**
	(1.023)	(2.463)	(0.691)	(1.268)	(1.026)	(1.509)	(1.023)
C. Labor Supply							
8. Hours Online	0.089**	0.122**	0.034	0.124***	0.090***	0.089***	0.089**
	(0.034)	(0.056)	(0.025)	(0.046)	(0.034)	(0.034)	(0.034)
9. Work Days	0.033***	0.069***	0.015	0.038***	0.036***	0.050***	0.033***
	(0.012)	(0.026)	(0.011)	(0.015)	(0.012)	(0.015)	(0.012)
10. Attrition	-0.007**	-0.010	-0.007*	0.003	-0.009***	-0.013***	-0.007**
	(0.003)	(0.007)	(0.004)	(0.004)	(0.002)	(0.005)	(0.003)
D. Wage Rate and Profit							
11. Income	0.113***	0.162***	0.083**	0.131***	0.118***	0.136***	0.113***
	(0.031)	(0.062)	(0.033)	(0.039)	(0.031)	(0.036)	(0.031)
12. Wage rate (no pen.)	0.051***	0.020	0.105***	0.038*	0.052***	0.051***	0.051***
	(0.015)	(0.023)	(0.015)	(0.019)	(0.015)	(0.015)	(0.015)
13. Wage rate (w/ pen.)	0.050***	0.0185	0.104***	0.032	0.051***	0.050***	0.050***
	(0.015)	(0.024)	(0.015)	(0.020)	(0.015)	(0.015)	(0.015)
14. Profit	0.113***	0.166***	0.081**	0.128***	0.118***	0.133***	0.113***
	(0.031)	(0.064)	(0.031)	(0.039)	(0.031)	(0.037)	(0.031)
N (Obs.)	480,070	207,084	272,986	427,612	184,450	343,733	480,070
Unique Drivers	43,605	27,777	15,828	40,821	3,397	34,965	43,605

Notes: Standard errors in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01. Column definitions: (1) Baseline specification using full panel. (2) Sample restricted to Gig workers only. (3) Sample restricted to Team workers only. (4) Excludes the first four months of tenure. (5) Restricted to workers present both pre- and post-intervention. (6) Restricted to observations with hours data. (7) Treatment effects with distance control included.

B Impact of AI Tool on Entry of DHH Workers

This section examines how the introduction of the AI tool affects the entry of DHH workers onto the Platform. Our analysis in the main text is based on workers who join the Platform and complete at least one order. However, if the AI tool makes work on the Platform more attractive for DHH workers, it could also spur more DHH workers to join.

We cannot observe the set of DHH workers potentially interested in working on the Platform. Instead, we collapse the data by disability type (DHH or non-diabled) and year-week. We have 3 years of data, so 3*52 = 156 year-weeks, and there are two disability types, so we have 312 observations. We then estimate a model very similar to equation (3):

$$Y_{gt} = \alpha + \gamma \cdot DHH_g \times Post_t + DHH_g + \xi_t + \varepsilon_{gt}. \tag{B.1}$$

where g denotes group (i.e., DHH or non-disabled) and t denotes time in year-weeks; Y_{gt} is the inverse hyperbolic sine of the number of new entrants on the Platform in group g during year-week t; and DHH_g is a dummy for being DHH. We are interested in estimates of γ.

Table B.1: Impact of AI Tool on Entry of DHH Workers: DiD Results

	(1)
$DHH_g \times Post_t$	0.261***
	(0.110)
Observations	312
R-squared	0.924

Notes: An observation is a year-week × disability type (i.e., DHH or non-disabled). Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.1 presents the regression results. The AI tool increases the number of DHH registrations on the Platform by 26%. This is a sizable recruitment effect.

C Qualitative Data

C.1 Interviews

After conducting a preliminary analysis of the data on measuring the productivity of DHH workers, we sought to better understand their daily routines, how they work, and the challenges they encounter. Therefore, we conducted interviews with seven DHH workers. To facilitate communication with DHH workers, we hired a translator proficient in sign language.

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

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

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. 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.

C.2 Fieldwork

To deepen our understanding of the working conditions and everyday challenges faced by DHH 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.1). This allowed us to better understand the initial challenges faced by novice workers. Second, another author spent several days directly observing DHH 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.1). 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. All photographs were taken by the author group. Images were converted to black-and-white to anonymize the color of the Platform's uniforms, and all Platform-identifying information was removed to preserve anonymity. Facial images of the authors are retained, whereas facial images of workers are obscured.

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.1: Immersive Fieldwork with Delivery Workers



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



(b) Author observing a DHH 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.2 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.

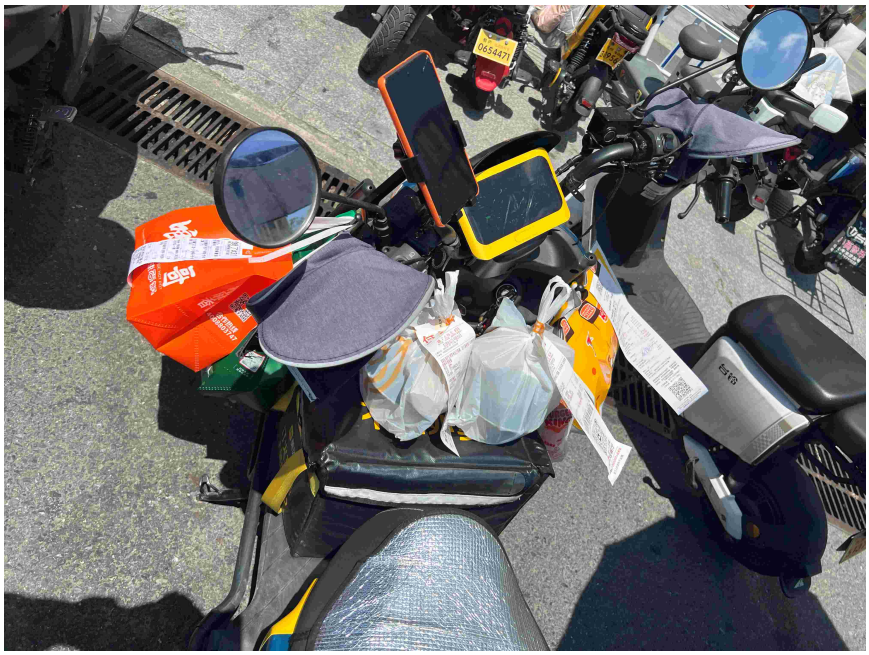
- To avoid being assigned too many unwanted orders by the station manager or platform, workers may delay clicking 'confirm delivery.' For example, when we delivered to Westlake University in the afternoon, the worker didn't want to take any more distant orders, so he delayed confirming the delivery.
- 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.2: Essential Equipment for Outdoor Work



Notes: Figure C.2 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.

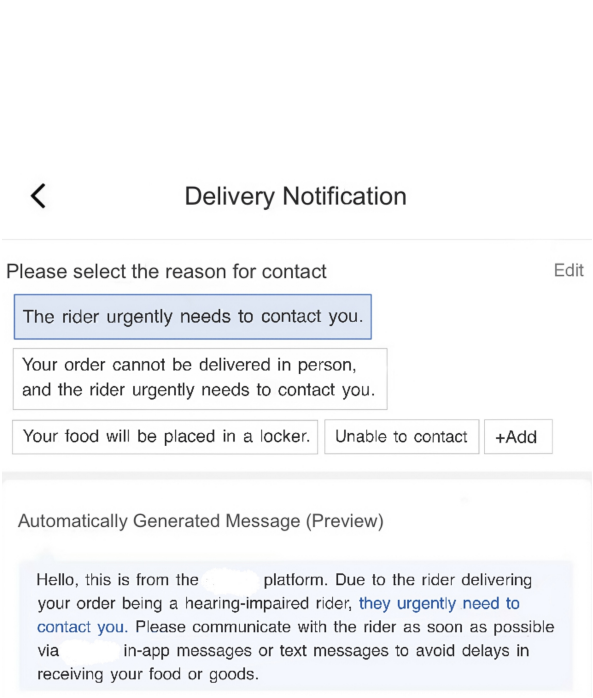
Figure C.3: Concurrent Orders During Peak Hours



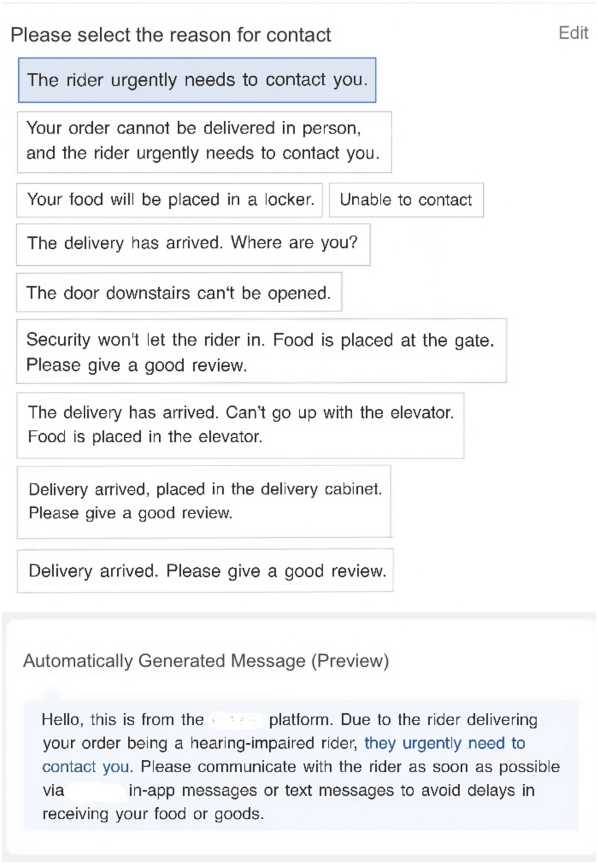
Notes: Figure C.3 illustrates multiple orders being simultaneously handled by a delivery worker during peak hours. Experienced workers typically manage between 6 and 10 concurrent orders during peak hours. The maximum number of orders a rider can handle simultaneously is set by the platform's algorithm, considering factors like tenure and past performance. Further details on the assignment algorithm are discussed in Section 1.

D Details on the AI Tool: 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 (IOC) 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 IOC 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 IOC 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, the IOC 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.