Empowering Inclusive Work*

Yanyou Chen
University of Toronto

Mitchell Hoffman
UC Santa Barbara and NBER

Huilan Xu
Zhejiang University

Zhe Yuan
Zhejiang University

April 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 calling system designed to improve customer communication for DHH workers. Using a difference-in-differences design comparing DHH and non-disabled workers before and after the AI tool, we find that AI increases the speed and productivity of DHH workers, especially on tasks involving customer interaction; substantially reduces negative customer reviews; and increases labor supply and retention. AI eliminates one-third of the disability hourly pay gap and significantly increases the profitability of DHH workers for the platform.

Keywords: AI, productivity, disability, personnel economics, gig work, deaf or hard of hearing
JEL Codes: M50, J14

*We thank Amanda Agan, David Autor, Heski Bar-Isaac, Gordon Dahl, Chiara Farronato, Peter Kuhn, Xiang Hui, Rafael Lalive, Imke Reimers, and numerous seminar participants for helpful comments. We also thank Xiaonan Zhu, Weiyue Zhang, Ruibo Sun, Yaqi Du, Yu Xu, 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. People with different disabilities face different challenges and may be affected by AI in different ways. Focusing on a single, common disability allows us to more clearly examine the relevant challenges. Moreover, hearing impairment is a disability that gig platforms around the world increasingly seek to accommodate, yet important challenges remain—making our findings broadly relevant.[2]

Our analysis draws on a rich administrative dataset from the Platform (Section 1), covering 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 also includes labor supply variables on both the intensive and extensive margins, as well as information pertaining to worker compensation and Platform profits. In the middle of our observation period in mid 2022, the Platform introduces an AI-based "Intelligent Outbound Calling" feature designed 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 true on the final segment of trips, from restaurant pickup to customer delivery—the stage involving interaction with customers—where disabled workers are 4% slower. DHH workers are substantially more likely to experience "left tail" outcomes, having 9% more late deliveries and 31% more bad customer ratings (with the caveat that bad customer ratings are rare). Despite this, pre-AI, DHH workers have substantially higher labor supply: they work 5% more hours per week and are 41% less likely to quit. The higher labor supply of DHH workers more than offsets their small efficiency gap, with DHH workers completing more orders per week and generating higher weekly profits for the Platform than non-disabled workers. However, because DHH workers work slower and get more bad ratings, they earn 10% less per hour than non-disabled workers, similar to disability wage gaps in

[2]About 8% of Chinese adults have moderate to severe hearing loss (Wang et al., 2025). China's rate is broadly in line with the World Health Organization (2025) global prevalence rate of 5% for disabling hearing loss (i.e., hearing loss of more than 35 decibels). Chen et al. (2024a) report that there are nearly 100,000 registered DHH couriers in China. Rates of US hearing impairment vary, from 3.7% of all Americans (Erickson et al., 2026) to 13% of US adults with some impairment (Madans et al., 2021).

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 calls sound natural. Rolled out suddenly to all DHH workers—but not to non-disabled workers—we exploit this as a natural experiment to estimate, via difference-in-differences, how the AI tool affects workplace outcomes for DHH workers. The AI tool causes DHH performance to substantially improve. DHH workers' speed of work increases, though only on the trip segment involving customer communication where they lagged the most previously. They also have far fewer late deliveries and bad ratings, with AI fully eliminating the pre-AI gap in late deliveries and cutting the gap in bad ratings by two-thirds. 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 higher profitability of DHH workers, with Platform gains far larger than costs. AI causes the disabled worker pay gap to fall from 10% to under 7%, a decrease of one-third. While DHH workers with non-profound (i.e., less severe) hearing loss also benefits, gains are largest among those with profound hearing loss. Overall, the results support the key mechanism for effects is improved communication with customers, as opposed to alternatives like increasing customer sympathy for DHH workers. We also find larger gains for DHH workers in stronger local labor markets.

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. Exploiting changes to workers compensation programs, Aizawa et al. (2025) find that disability accommodations by firms increase disabled worker earnings and employment.

[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. As surveyed by Fruits and Stout (2026), work examines how AI affects outcomes for different types of workers, including accountants (Choi and Xie, 2025), translators (Merali, 2024), 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 workers and reduce inequality (Autor, 2024; Acemoglu et al., 2026). 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).

[5]Relative to gender and racial gaps, there are many fewer studies on disability wage gaps. We discuss recent work below in the context of our results.

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 US apps like Uber Eats or DoorDash but on a much larger scale. Together, the two main Chinese platforms engage approximately 10 million delivery 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 requirements: be at least 18 years of age, possess valid identification, and have no disqualifying criminal record. These standards are uniform across applicants, with no separate criteria for workers with disabilities.

Disability status on the Platform. Across both major Chinese food delivery platforms, most disabled workers are DHH, or *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), with further details on Chinese disability levels in Appendix E:

- Level 1, which is hearing loss of greater than 90 decibels, abbreviated as $>$ 90 dB HL. This is profound hearing loss, where someone is unable to hear even a shouted voice

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

and spoken speech is distorted (Olusanya et al., 2019). Extremely serious barriers occur regarding participation in social life (GB/T 26341-2010).

- Level 2 or 81-90 dB HL. This is severe hearing loss.
- Level 3 or 61-80 dB HL. This is moderately severe hearing loss.
- Level 4 or 41-60 dB HL. This is moderate hearing loss.

Based on our qualitative work and understanding of the Platform, essentially all certified DHH workers disclose their disability to the Platform, a point we discuss further in Appendix F. Drivers know their level since it is listed on their certificate of disability.

Both before and after the introduction of the AI tool, DHH couriers are marked with a badge in the platform interface visible to customers. However, customers may not always attend to this information. At the same time, there is a general expectation in China that drivers call customers upon arrival, and this can pose communication challenges for DHH workers.

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.

Work model is also referred to as registration type. If a worker wants to switch work models, they need to re-register. Based on our conversations with the Platform, we believe that relatively few workers switched work models during the period of our data.

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.

[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.

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. The algorithm's formula does not include disability status as an explicit input.[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.

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 $r \cdot \frac{\beta}{\beta-1}(1-h)^{\frac{\beta-1}{\beta}}$, 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 expres-

[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.

sion $h = 1 - r^{\beta}(wz)^{-\beta}$. 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 DHH workers about work before the AI tool, namely, that there were often significant bottlenecks toward making quick deliveries.[9]

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

[9]Disability scholarship (e.g., Krieger, 2010; Lane, 1992) emphasizes that differences between disabled and non-disabled people reflect both individual traits and environmental factors (e.g., discrimination, accommodations). Thus, A2 is a joint statement about capabilities and the environment, not inherent ability.

workers to increase labor supply, and thus two forces together work toward increasing total output and profits.

Framework extension: multiple severities of disability. While the framework has one severity of disability, it is easily extended to have multiple severities. Suppose, as in our setting, that there are workers who are profoundly disabled ("PD") and those who are non-profoundly disabled ("NPD"). Extending the logic of A1 and A2, it is natural to assume that $z_{PD} < z_{NPD}$ and $r_{PD} < r_{NPD}$. These assumptions yield pre-AI implications mirroring those between disabled and non-disabled. Turning to the impact of the AI tool, suppose communication situations vary in difficulty. Workers with non-profound hearing loss may be challenged primarily in difficult communication situations, whereas workers with profound hearing loss may face challenges in "simpler" interactions. If the AI tool helps resolve simpler communication situations but not the most difficult ones (Autor, 2015), then productivity gains from the AI tool will be larger for profoundly disabled workers.

Implication 5. Before the new AI tool, profoundly disabled workers have lower efficiency than non-profoundly disabled workers, but differences in labor supply, output, and profits are ambiguous. The AI tool will have larger effects on profoundly disabled workers relative to non-profoundly disabled workers.

Framework extension: multiple local labor markets. Since the Platform operates across China, we observe AI impacts across a range of local labor markets, from the "rust belt" of Northeast China to the economically more prosperous provinces in the Southeast. Disabled workers likely have stronger outside options when the local labor market is stronger, which can be proxied by a lower local unemployment rate, i.e., $r_D^S > r_D^W$, where r_D^S and r_D^W denote disabled workers' outside options in strong and weak local labor markets, respectively. Since $\frac{\partial^2 h}{\partial r \, \partial z} > 0$, an AI-induced increase in productivity leads to larger increases in hours for disabled workers in stronger local labor markets than for disabled workers in weaker local labor markets.

Implication 6. The AI tool will have larger effects on hours, output, and income for disabled workers in stronger local labor markets compared to disabled workers in weaker local labor markets.

Framework simplicity. This framework is deliberately simple and abstracts from several features of reality. First, true worker "productivity" is multidimensional—encompassing speed at different stages of delivery, late deliveries, and bad ratings—whereas the model summarizes these dimensions with a single parameter z. Second, the model does not explicitly distinguish between hours worked and retention. In our setting, it is useful to think of

"labor supply" broadly as including hours on the Platform, timing of work, and continued participation, e.g., workers with worse outside options may both work more hours and be less likely to quit the Platform. The empirical analysis examines these margins separately and documents pre-AI differences between DHH and non-disabled workers, as well as how the AI tool affects those differences.

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. Drivers with multiple disabilities were excluded from the panel shared with us. In total, the data contain hundreds of thousands of worker-week observations.

The panel allows us to track each worker from the start of their tenure on the Platform and to analyze how their performance evolves over time. By focusing on newly registered workers, we compare DHH and non-disabled workers from their point of entry, avoiding stock-sampling concerns related to differential tenure.

The AI tool was rolled out suddenly at t_0 in mid 2022 to all DHH workers. To avoid identifying the particular Platform, we do not specify the exact month of t_0, but it is a date during April 2022-September 2022, thus giving 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 (or quantity) 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, via direct assignment by the Platform or via acceptance by the worker. The time for this first step is called the *order acceptance time*. Second, once the worker receives the order, they pick it up from the restaurant. The waiting time for food

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

preparation is included in this step, which concludes when the worker actually receives the food from the restaurant. The time for this second step is called the *pickup time*. Lastly, the worker delivers the order from the store to the customer, the time for this being the *delivery time*. 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, customers on our Platform click a button to indicate a Bad Rating, a Good Rating, or nothing (most common). 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). Drivers incur a monetary penalty from the Platform for each bad rating received, giving bad ratings direct financial consequences.

Drivers can also receive Good Ratings, but we emphasize these much less, as they have no effect on worker compensation or order assignment, and because the Platform emphasizes them much less. We therefore report results for Good Ratings only in the Appendix.

We cannot analyze accidents or injuries, as the Platform does not comprehensively record them in the administrative data we received.

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 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 exits the Platform 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

[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.

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 the per-order profit times the number of orders completed, 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 400 CNY.[13] 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 retention and reputational spillovers, net of the penalty paid by drivers to the Platform.

Control variables. The data also include rich control variables, detailed below in Section 2.1, 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 2024, this includes:

1. Interviews conducted by coauthors with 7 DHH workers to learn about their daily routines, challenges, and communication methods.

2. On-the-ground observations during which two coauthors accompanied DHH workers and delivered orders full-time alongside them. Coauthors completed slightly more than 100 deliveries over multiple days. The purpose of these observations was to gain

[13]The biggest determinant of this is a bonus the Platform pays to workers after their first week on the Platform.

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 D 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

Sample restriction. 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 shows the distribution of disability severity among DHH workers, with 54% having profound hearing loss (Level 1 in the Chinese system).[14]

Panel D of Table 1 compares DHH and non-disabled workers in terms of geography and distance variables. DHH workers are slightly more likely to be located in big cities than non-disabled workers.[15] DHH workers' orders have slightly higher travel distance than non-disabled workers.

Table 2 provides a detailed summary of all outcome variables. Starting with Total Output in Panel A, workers complete an average of 152 deliveries per week, with a median of 127. The standard deviation is large at 142, indicating considerable variation in total deliveries. Peak-hour orders, which occur during high-demand times, account for about 91 deliveries per worker-week, indicating 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

[14]The share of drivers with profound hearing loss is higher in our sample than the overall Chinese population. A natural explanation is lower barriers to entry in gig work (e.g., no interview requirement) for people who are profoundly DHH, rather than differential disclosure to the Platform. Appendix F discusses further.

[15]One explanation is greater opportunities for DHH workers in large cities. Another is the historical concentration of specialized deaf schools there. We do not observe schooling histories.

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 9.6, 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 bad ratings.

Workers spend an average of 40 hours online per week. 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.[16]

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

Appendix Table A.1 presents summary statistics separately for gig and team workers. There is less variability in outcomes within each work model separately.[17]

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:

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

[16] While hours on the Platform are higher than for US Uber drivers (Cook et al., 2021), they are very typical for hours in middle-income and developing countries, where gig work often serves as a primary, full-time source of income (Chen et al., 2024b; Adhvaryu et al., 2026).

[17] For example, consider hours worked. Among all workers, the standard deviation is 29 hours. Among gig and team workers separately, the standard deviations are 22 and 25 hours, respectively.

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. We include year-week fixed effects (capturing common shocks during the outcome period), city-of-work fixed effects (roughly 300 cities), age fixed effects, gender, team status (gig vs. team), and registration year-month fixed effects (capturing cohort differences at entry). In our main specifications, we do not control for travel distance per order—a measure of task difficulty—or for worker tenure, as these variables represent mechanisms by which DHH and non-disabled workers may differ.[18] In the Appendix, we show that controlling for travel 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.

In addition to overall differences, we also present outcome-tenure curves separately for DHH and non-disabled workers. This allows us to analyze 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. To do this, we regress outcomes on interactions between DHH status and worker tenure month indicators, controlling for the same controls listed in Equation (2). The excluded group is non-disabled workers in month 1, and we plot the sum of the regression coefficients with the excluded group mean. Outcome-tenure curves are similar with no controls.

Functional form of dependent variable. For each variable besides late deliveries, 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 sample to weeks with positive orders, most of our outcome variables are almost never equal to zero, and we avoid the "log of zero"-like problem with inverse hyperbolic sine (Chen and Roth, 2023). In line with Chen and Roth (2023), we estimate non-transformed OLS models for late deliveries and bad ratings, and divide by a pre-AI mean to calculate percentage differences. Late deliveries is analyzed in terms of late deliveries per 10,000 orders, whereas bad ratings is a weekly count. For worker attrition, which is binary, we analyze linear probability models.

[18] 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.

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. Our findings are generally similar without control variables, as can be seen Figure A.10, which plots raw means over time for several outcomes for DHH and non-disabled workers.

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 3.4% 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 between 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 4.5% longer, which is about one minute more. Overall, DHH workers do 1.9% fewer orders per service hour than non-disabled workers.[19]

Figure 2 presents the tenure-productivity relationship for each of the three order stages. Time spent in each stage declines with tenure for both groups, consistent with learning by doing (Shaw and Lazear, 2008).[20] For the first stage, order acceptance, differences are present in the first three months of tenure, but the two lines converge starting in the fourth month. In contrast, for the final stage, customer delivery, the gap is more persistent. That is, not only is the DHH difference largest for delivery time, but it also takes much longer to converge, and the average remains higher for DHH workers throughout.

The time numbers on each stage of the order fulfillment 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 9% 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 deliveries per 10,000 orders.

[19] Orders per service hour is calculated as $\frac{60}{t_1+t_2+t_3}$, where the denominator is the sum of average minutes spent in each stage of an order. This measure doesn't use our hours data, which capture time available for work rather than time completing orders. Because drivers can handle multiple orders concurrently, the sum of average stage times can be less than actual clock time spent delivering orders, so orders per service hour should be interpreted as a processing-rate measure rather than literal orders per hour doing deliveries.

[20] This pattern is also consistent with differential attrition, which we discuss in Section 3.5. The main pattern of order acceptance, pickup, and delivery time decreasing with tenure is also present in balanced panels.

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. Our measure of work quality is bad customer reviews. Figure 1 shows that DHH workers receive 31% 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, results on speed and bad ratings support Assumption 2 in Section 1.1 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 4.9% more hours per week than non-disabled workers, conditional on controls, equivalent to an extra 2 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 4.3% days per week than non-disabled workers, equivalent to an extra 0.2 days per week.

Beyond working more hours and days, Figure 1 shows that DHH workers are 41% 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 3-4 times 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 5.3% 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. The tenure-productivity curves in Figure 6 are obtained from the same specification used to construct Figure 2. As with the speed variables, performance increases with tenure. However, for total output, DHH and non-disabled workers have very similar numbers of orders at a given tenure level. The advantage of DHH workers therefore arises not from higher output conditional on tenure, but from their having longer tenure and lower quit rates. In other words, DHH workers tend to operate at higher tenure levels of the learning curve than non-disabled workers.

The output gap is largest in the team model. Among team workers, DHH workers complete 8.7% more orders per week than non-disabled workers, about 18 additional orders. Among gig drivers, the difference is 0.7% 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 5.4% more in weekly gross income relative to non-disabled workers.[21] 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 9.4% lower hourly wages. When penalties from bad ratings are incorporated, the hourly wage gap widens slightly to 10.2%. These estimates are broadly 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 one-tenth of the 10% wage gap. Most of the pay gap reflects DHH workers having similar income to non-disabled workers (in the sample with observed hours used to compute wages) despite working more hours.

Turning last to Platform profits, DHH workers generate 12% 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.

Heterogeneity in pre-AI differences by DHH severity. Figure A.2 presents heterogeneity in pre-AI differences in outcomes according to disability severity. We split by DHH workers with profound HL (Level 1 in Chinese system; 54% of the sample) and DHH workers with non-profound HL (Levels 2-4). Profound HL workers tend to obtain worse outcomes than non-profound HL workers, though not all differences are statistically significant.

Starting with efficiency, profound HL workers get 48% more bad ratings than non-disabled workers, whereas non-profound HL DHH workers get 14% more bad ratings than non-disabled workers. There is no difference between non-profound HL workers and non-disabled workers in late deliveries, whereas profound HL workers have 16% more late deliveries. Turning to labor supply, the overall DHH hours advantage is driven by non-profound HL workers, who work 17% more hours per week than non-disabled workers. Profound HL workers work 12% fewer hours per week than non-disabled workers. Differences in total output, income, and Platform profits are likewise driven by non-profound HL workers.

Both groups earn lower hourly wages than non-disabled workers, but, interestingly, the wage gap is larger for non-profound HL workers (13% gap vs. 7% gaps). This reflects their

[21] 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.

substantially higher hours. Profound HL workers may concentrate work in more lucrative days or times, consistent with greater communication challenges.

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 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 emphasized helping DHH workers as part of its broader corporate social responsibility efforts. At the same time, the tool had the potential to increase profits (e.g., by increasing productivity, reducing costly delivery errors).

How the tool works. Under the IOC feature, DHH workers make natural-sounding, automated calls to customers. 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 G.1 illustrates both the default prompts offered by the platform and examples of customized prompts commonly used by DHH workers.[22] After the DHH workers selects a prompt, the app initiates an automated voice call to the customer. Each call begins with an introductory message clearly stating that the caller is DHH (*wú shēng*) followed by a specific message conveying the worker's issue or request, e.g., "Your food will be placed in a locker". The tool is strictly text-to-voice instead of voice-to-text, i.e., customers can respond in the app instead of on the call.

How the tool uses AI. The IOC uses AI-based text-to-speech technology, which converts written text into natural-sounding speech, and is widely used in voice assistants, audio-

[22]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.

books, and accessibility services. Pre-AI generated speech can perform basic functions, but often lacks naturalness and emotion. AI has brought significant breakthroughs to this area. Deep learning-based models like WaveNet and Tacotron use neural networks to directly synthesize speech waveforms, significantly improving fluency and making prosody, intonation, and pauses more human-like. AI voices are more pleasing to customers than traditional robocalls (Shah et al., 2023; Arcot, 2025). AI also enables rapid multilingual adaptation and diverse style simulation, capable of replicating specific voice characteristics and expressing different emotions.

The advantages of the IOC tool include: first, enhanced customer responsiveness, as customers are significantly more responsive to phone calls compared to app notifications, and voice communication is perceived by the Platform as "warmer" than text, ensuring prompt and higher-quality interactions; 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 announced the AI-based tool to DHH drivers through multiple in-app notifications, making awareness of the tool widespread. Our administrative data do not record individual-level usage. However, the Platform informs us that, to their knowledge, all DHH workers use the AI tool and that the tool is popular. This is validated by our qualitative evidence: all 10 DHH workers whom we interviewed or worked alongside reported using the AI tool.

3.2 Overall Effects of the AI Tool

Estimation strategy and graphical presentation of results. To assess the impact 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:

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

where $Post_t$ is a dummy equal to 1 if the current week is on or after the introduction of the AI tool at time t_0 in mid 2022. γ captures the differential change in outcomes for DHH workers after the introduction of the AI tool relative to non-disabled workers. Under the parallel trends assumption, γ identifies the average treatment effect on DHH workers. Worker effects λ_i capture fixed differences across workers while time effects δ_t capture common shocks. We

present our baseline DiD results without time-varying controls. 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 underlying equation (3) is that, absent the introduction of the AI tool, outcomes for DHH and non-disabled workers would have evolved similarly over time, conditional on controls. 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 concerns about differential trends using event studies.

Figure 3 summarizes DiD estimates of the AI tool's impact across all outcomes. Each bar shows an estimate of γ from equation (3) and represents the percentage change in 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 1.6%. 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. Delivery time decreases by 3.5% for DHH workers, closing roughly three-quarters of the pre-AI gap of 4.5%. In contrast, effects on order acceptance time and pickup time are economically negligible and statistically insignificant.

In terms of magnitudes, the reduction in delivery time of 3.5% is equivalent to moving about 3 months down the delivery time tenure curve.

Panel (b) of Figure 4 shows event study estimates of the impact of the AI tool on customer delivery time (i.e., the final stage). There are not systematic pre-trends, consistent with the identifying assumption in DiD. In the first couple months of the AI tool, delivery time may actually increase (though not significantly so), consistent with drivers taking a little time to use the AI tool effectively (Brynjolfsson et al., 2021). Effects on delivery time gradually turn negative, with effects seeming to gradually increase in magnitude.

The impact of the AI tool on substantially late deliveries is even larger, with a 11% reduction for DHH workers. The IOC entirely closes the late delivery gap between DHH and non-disabled workers. This is seen in raw means in panels (e) and (f) of Figure A.10.

As seen in Appendix Figure A.3, reductions in delivery time are pronounced and economically similar for both gig workers and team workers. We also see reductions in late deliveries for both types of workers, but the reductions are more pronounced for team workers. This could be because gig workers take many more orders after the IOC and are therefore more

likely to experience late deliveries, for example when carrying more orders simultaneously during peak hours, which increases the likelihood of delivering late.

Productivity variables: quality. Figure 3 shows that the AI tool has an especially large effect on bad ratings, reducing bad ratings by 22%. This closes roughly two-thirds of the pre-AI gap of 31%.[23]

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.

Panel (c) of Figure 4 shows effects on bad ratings over time in an event study. As for delivery time (in panel (b)), there is no pre-trend. Effects appear to grow somewhat over time, potentially reflecting adjustment costs in learning about the AI tool, though confidence intervals are larger.

The productivity results, both on quantity and quality variables, suggest that the AI tool alleviates frictions at the delivery stage: driver speed at earlier stages of the delivery process is unchanged, while delivery is faster, and customer satisfaction improves. 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.[24]

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 tool increases DHH weekly hours online by 10.6%, a tripling of the pre-AI gap of 4.9%. The AI tool increases work days per week by 3.5%, 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 21%, further expanding the pre-AI gap of 41% by half. 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

[23]The sizable negative impact of the AI tool on bad ratings is robust across many methods, which is important given that bad ratings are rare. For example, AI reduces the chance of receiving any bad ratings in a week by around 20%.

[24]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.

retention are present in both gig and team workers, as seen in Appendix Figure A.3. Overall survival of both groups is somewhat lower in the post-period, reflecting a stronger labor market in China in 2023 than 2021, but retention improves differentially for DHH workers.

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 workers have 5% 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 15%, 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 11% 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 of 3.6%. These estimates imply that roughly one-third 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 10%, these gains imply that IOC closes one-third 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.[25]

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 17%, a sizable and statistically significant effect. Because the tool was developed internally

[25]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. Gentry (2025) finds that gaining access to Uber closes less than 2% of the disability pay gap.

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 Benefits and Costs to the Platform

The Platform's average profit per order is confidential. However, assuming a profit per order of 7 CNY—about $1 USD—the 17% improvement in DHH worker profitability works out to roughly $58.5k USD per month as a benefit from the AI tool.[26]

The Platform's cost of the tool consists primarily of engineer time during development, namely the off-hours work of three engineers over several months. We conservatively assume that each engineer spent 10 hours per week for three months, implying a total time investment of 390 hours—less than 0.2 annual FTE—and thus no more than about $20,000 USD in labor cost. Under a simple comparison of benefits and costs, the AI tool pays for itself in less than a month: specifically, in 20/58.5 months, or about 1.5 weeks.

In calculating the increase in profitability for DHH workers of 17%, this implicitly assumes that additional orders completed by DHH workers do not displace orders from non-DHH workers, corresponding to an infinitely elastic labor demand curve at the platform level. To relax this assumption, suppose that a fraction δ of additional orders completed by DHH workers are displaced from non-DHH workers. In this case, the net increase in platform profits is scaled by $(1 - \delta)$. Even under substantial displacement—for example, $\delta = 0.5$, implying that half of additional DHH orders come at the expense of non-DHH workers—the tool would still generate roughly $29k USD per month in additional profits and pay for itself in under one month. Even with $\delta = 0.75$, the tool would still pay for itself in under two months.

3.4 Heterogeneous Treatment Effects

A central question regarding the deployment of AI in the workplace is how treatment effects vary across different workforce segments. For example, Brynjolfsson et al. (2025) show that

[26]There are roughly 1,000 DHH workers on the Platform at any given time. Assuming 600 orders per month, we have 1,000*600*1*1.17*(1/12) = $ 58500 USD per month.

generative AI tends to confer larger benefits to novice or lower-skilled workers. We focus our analysis on four key dimensions: pre-AI tenure, pre-AI productivity, disability severity, and local labor market conditions. For each dimension of heterogeneity, we split the sample based on the median, and estimate simple interaction term models.

Pre-AI tenure and productivity. We find no consistent evidence of heterogeneous treatment effects based on pre-AI tenure (Figure A.11) or pre-AI productivity (Figure A.12), with pre-AI productivity measured in orders per week. In each panel, there is only statistical significance for 1 out of 7 outcomes in terms of heterogeneity. That is, the AI tool's effectiveness does not systematically vary with a driver's tenure or average productivity.

Disability severity. We observe substantial heterogeneity in effects based on severity of disability. Figure 9 plots the coefficients for workers with profound hearing loss (meaning Level 1, hearing loss of more than 90 decibels) versus those with non-profound hearing loss (Levels 2–4), whom we refer to as profound HL and non-profound HL workers.[27] Profound HL workers, who are 54% of DHH workers in our sample, cannot hear even shouted speech.

We observe substantial heterogeneity in effects based on severity of disability, usually in the direction that profound HL workers benefit more. Orders per week increase by 6% for non-profound HL workers, but by 16% for profoundly impaired, with a p-value of 0.08 on the difference. Bad ratings decrease by 7% for non-profound HL workers, but decrease by 40% for profound HL workers ($p = 0.02$). As seen in Figure A.2, pre-AI, the bad rating gap is largest for profound HL workers (and profound HL workers do 15% fewer orders than non-profound HL workers), so the AI tool also helps close the pre-AI gap by disability severity within DHH workers. This suggests that the AI tool is most effective at reducing communication frictions for those who face the steepest barriers.[28]

Relative to non-profound HL workers, profound HL workers also see substantially greater gains in income ($p = 0.08$) and wages ($p = 0.05$). While statistical significance is sometimes borderline on differences, the differences are economically quite sizable, e.g., wages go up by around 1.5% for non-profound but by 7% for profound HL workers.

For attrition, profound and non-profound HL drivers show similar effects. One potential explanation is that profound HL workers have fewer outside labor market options regardless of the tool, whereas the AI tool effectively tips the scales for non-profound HL who might

[27] Unlike our other heterogeneity dimensions, severity is defined only within DHH workers, so an indicator Profound_i for profound HL is absorbed by worker fixed effects λ_i and the standard four-term interaction specification reduces to two terms: $DHH_i \times Post_t \times \text{Profound}_i$ and $DHH_i \times Post_t$. The worker fixed effects also address the standard concern that Profound_i may be correlated with unobservables.

[28] Interestingly, non-profound HL workers exhibit larger improvements in delivery time and orders per service hour. One possible explanation is that profound HL workers respond more along the extensive margin—working more and completing more orders—so that their gains are reflected more in total output than in per-order speed.

otherwise have exited.

We focus on splitting by profound and non-profound HL both because this is roughly a 50/50 split (and thus good for statistical power) and because profound HL can be qualitatively different from non-profound HL. However, results are broadly similar when we estimate dosage-response heterogeneous treatment effects, as seen in Appendix Figure A.13.

Local labor market conditions. Figure A.14 examines heterogeneity in AI effects by local labor market conditions. Exploiting that our Platform operates nationwide, we split China's 31 provincial-level administrative units into higher and lower unemployment provinces, where the 2021 average unemployment rates were 3.6% and 2.5% respectively.

On the 12 variables where we detect significant positive AI effects (i.e., excluding order acceptance and pickup time), in 7 of 12 variables, we observe a significantly more beneficial effect in lower unemployment rate provinces. For example, AI has no effect on late deliveries in higher unemployment provinces, while AI reduces DHH late deliveries by 14% in lower unemployment provinces. Likewise, there is no AI effect on DHH hours online, work days, or orders in higher unemployment provinces, but sizable improvements in all these dimensions in lower unemployment provinces. While DHH driver income increases by 8% in higher unemployment provinces, it increases by 17% in lower unemployment provinces.

3.5 Threats to Identification and Robustness Checks

Existence of concurrent policy changes. Our event study graphs show no clear pre-trends prior to the AI tool's introduction. A threat to identification would be if the Platform made other policy changes around the same time as the AI tool that differentially affected DHH workers. According to the Platform, there were no changes made close to the introduction of the AI tool, and in fact, no meaningful other changes for DHH workers over the sample period.

Several of our quantitative results are strongly consistent with this. Suppose there was another policy launched by the Platform (or another entity like the Chinese government) to improve the productivity of DHH workers. Such a policy would likely show effects on all three stages of the delivery process: order acceptance, order pickup, and order delivery. However, we only observe effects of the AI on speed of order delivery, which was the only segment that the AI tool is designed to help with, as drivers do not use the tool for communicating with restaurants or for accepting orders.

In addition, the observed heterogeneity by disability severity speaks to the plausibility of the identification assumption. If there was an alternative Platform intervention that benefited DHH workers, it would not necessarily bestow greater benefit on profoundly impaired

DHH drivers relative to non-profoundly impaired DHH drivers.

Accounting for the transition period. Like many new technologies, usage of the AI tool was not instantaneous at 100% usage. Rather, according to the Platform, there was a transition period of a few months where DHH drivers learned about the new AI tool and how to best use it. This is supported by the event study graphs, which often show impacts that are more muted in the first few months, but that increase over time. Since we are primarily interested in the longer-run effects of the AI tool, we also re-do our DiD analyses 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 and often become larger. The reduction in bad ratings intensifies from 22% to 25%, the increase in output per hour grows from 1.6% to 2.1%, the increase in weekly orders grows from 11% to 13%, and the increase in income grows from 11% to 13%. Thus, our baseline DiD estimates, which include the first several months of the AI tool being available, may represent a more conservative lower bound of the tool's longer-run effects.

Improvements to DHH workers or declines for non-disabled workers? Could our DiD estimates be driven by non-disabled outcomes getting worse instead of DHH outcomes getting better? Several pieces of evidence speak against this. First, raw means of various outcomes (Appendix Figure A.10) show substantial improvements for DHH workers in levels. Second, non-disabled workers getting worse would not generate the observed heterogeneity by disability severity, with larger AI benefits on more severely disabled workers.

One particular form of non-disabled workers getting worse would be if DHH workers take delivery opportunities away from non-disabled workers. This is unlikely to be quantitatively important for non-disabled worker outcomes because the share of DHH workers on the Platform is small (i.e., less than 1% of workers on the Platform).

COVID. Could our results be driven by COVID lockdown policies in China? Chinese cities instituted lockdown at different times at various periods in 2021 and 2022. We re-did our main DiD results excluding the period of strongest lockdown, and results are very similar. In addition, we have run our DiD results controlling for a measure of local COVID cases, and this has no effect on the results.

Differential attrition. Since the AI tool increases DHH worker retention, a concern is that this could "bias" our estimates of the effect of the AI tool on DHH non-attrition outcomes like productivity variables, hours, and wages. Given that productivity increases with tenure (see Figures 2, 6, and 7), if AI boosts DHH worker retention, then DHH workers will be more likely to be in higher-productivity parts of the learning-by-doing curve.

If we take the perspective of the Platform, our baseline DiD estimates should not be

"corrected" for possible differential attrition, as it is an important *mechanism* by which the AI tool affects outcomes. The Platform cares about the overall effect of the AI tool on outcomes, not the impact of the AI tool on particular workers.

At the same time, we are still interested in how AI affects outcomes separate from affecting retention, and such estimates might be more relevant for DHH workers themselves (i.e., how much will the AI tool affect my productivity). To shed light on this, and shut down differential attrition, we perform our DiD analysis restricting the sample 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 the balanced panel are quantitatively and qualitatively consistent with our baseline findings. We continue to observe significant improvements across key metrics: total weekly orders rise by 10%, weekly income increases by 10%, and bad ratings decrease by 20%. These strong effects are similar to our baseline estimates, indicating that differential attrition is not a central driver of the DiD estimates. Rather, the AI tool increases worker productivity directly in addition to increasing retention.[29]

Missing data on hours. As noted above, we have significant missing data on hours. Figure A.7 plots the missing rate by calendar time. Hours is missing for all workers before week 19 of 2021. In week 19 of 2021, it becomes available for gig workers, and then becomes available for team workers in around week 10 of 2022. There is very little difference between DHH and non-disabled workers in the frequency of missing records, indicating that missingness reflects centralized institutional practices rather than worker behavior or AI-driven interventions.

Because of the missing hours data, our estimates regarding hours-based outcomes, namely, hours online and wages, rely on a different sample than other outcomes.[30] To ensure that our constellation of results is not being driven by different samples, we redo our main results from Figure 3 while restricting the sample to observations where hours are non-missing. The results, presented in Figure A.4, demonstrate strong consistency with our main findings.[31] In the restricted sample, the AI tool drives significant improvements: orders per week increases by 13%, weekly income rises by 13%, and time spent between pickup and delivery decreases by 2.6%.

[29]Because the AI tool increases retention among DHH workers, the balanced-panel sample likely includes marginal DHH workers who would otherwise have exited. If these marginal workers are less productive on average, this compositional change would bias the estimated productivity effects downward. The positive and significant effects in the balanced panel should therefore be interpreted as conservative estimates, and likely a lower bound on the true productivity gains from the AI tool.

[30]Hours per service hour relies on the data on time spent on different parts of the production process.

[31]In the restricted sample, the AI tool has a negative effect on late deliveries and bad ratings, but the effect is no longer statistically significant, reflecting the smaller sample with hours data.

The role of late deliveries and bad ratings in load assignment. As noted in Section 1, late deliveries and bad ratings can affect the set of orders workers are assigned or able to select. Beyond direct productivity improvements, an alternative mechanism is that the Platform algorithm assigns treated workers easier or higher-quality orders as their performance metrics improve. In that case, observed gains in delivery speed, orders, hours, or income would reflect task reallocation rather than improved customer communication. This is a question about mechanisms rather than identification.

We address this concern in two ways. First, we re-estimate our main DiD specification while flexibly controlling for task difficulty, proxied by the distance between the restaurant and customer. As shown in Figure A.5, the estimated effects are very similar to the baseline results, indicating that changes in trip distance do not explain the observed DiD gains. Second, we conduct a mediation analysis by controlling for cumulative late deliveries and bad ratings to date. The estimated effects of the AI tool are largely unchanged, suggesting that improvements in these performance metrics account for little of the overall impact.

Overall, the results suggest that reductions in late deliveries or bad ratings, or the resulting reallocation of tasks, are unlikely to drive our other key DiD findings.

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 as much. The heterogeneity results by disability severity suggest this is unlikely. DHH workers with non-profound HL benefit substantially less than those with profound HL, suggesting that non-disabled workers might also benefit less, or possibly not at all. The Platform's decision not to roll out the AI tool to non-disabled workers is also consistent with the view that benefits for non-disabled workers would be smaller or negligible.

Adoption on one platform vs. both platforms. The other major food delivery platform in China introduces a similar AI tool for DHH workers later in the same year, after our Platform. Because the market is largely dominated by these two platforms, the period in which only our Platform offers the tool raises the possibility of cross-platform spillovers (e.g., worker or customer reallocation). To address this, we re-estimate our DiD results excluding the period in which the AI tool is available on our Platform but not on the other platform. The estimates are very similar. This suggests that our results are unlikely to be driven by the temporary presence of the tool on only one platform. The subsequent adoption of a similar tool by the other platform is also consistent with the tool being broadly valuable for DHH workers.

Reference dependence / income targeting. Could the pre-AI difference simply reflect that DHH workers target income, doing more hours to make up for lower efficiency?

This is inconsistent with our results by DHH severity, where profound HL workers have both lower efficiency and fewer hours than non-profound HL workers.

3.6 What the AI Tool Does Not Address: Qualitative Evidence

Although the AI tool substantially improves communication between DHH workers and customers, workplace outcome gaps remain. Qualitative evidence helps clarify the types of situations that allow such gaps to persist.

Our qualitative data were collected in summer 2023, more than a year after the AI tool was introduced. Consistent with what the Platform told us, in the workers we interviewed or worked alongside, all 10 of them used the AI tool. Thus, the remaining gaps do not reflect limited adoption of the tool.

Instead, the evidence points to communication frictions that the AI tool does not fully resolve. These include difficulty locating customers' exact positions, navigating delivery environments with complex physical layouts (such as condominium complexes with multiple entrances), and situations in which customers provide instructions that remain ambiguous without verbal clarification. A coauthor working alongside a DHH worker experienced the following, as recorded in his field notes, where "I" denotes the coauthor:

> The [DHH] 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.

A second situation arose while the same coauthor was working alongside the same DHH worker, illustrating cases in which written instructions remain unclear in the absence of verbal communication:

> One order was delivered to an urban village [chengzhongcun in Mandarin], but the address was inaccurate. Upon arriving, the worker used the intelligent outbound call to contact the customer, who messaged back, "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.

In these situations, the AI tool helps initiate contact when customers might otherwise not notice or respond to messages. However, ambiguity in customers' directions and physical delivery environments can persist—particularly in cases where clarification would normally rely on real-time verbal communication.

4 Conclusion

By 2050, roughly 10% of the world's population will have disabling hearing loss (World Health Organization, 2025). While DHH workers face intense labor market challenges including lower earnings and employment, research has struggled to identify policies or tools that can benefit DHH workers or disabled workers in general.

We comprehensively examine the relative workplace outcomes of DHH food-delivery workers in China, and analyze how DHH workers 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 greater profit for the Platform on a per-worker basis, they earn about 10% less per hour.

The introduction of a targeted non-generative AI tool on text-to-voice calling substantially enhances DHH worker performance. Work speed improves, driven entirely by faster completion of the final stage of order fulfillment—delivery to customers—where the AI tool is used. Orders rise and customer complaints fall. Retention, already higher for DHH workers, increases further. The AI tool decreases the disability wage gap by one-third—the largest impact on the disability wage that we are aware of any intervention for any disability—and also 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 illustrates 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.

Relevance for other disabilities. While our results pertain specifically to DHH workers, AI tools may also benefit people with other disabilities. Many human faculties involve prediction, the core function of modern AI (Agrawal et al., 2018). In the context of deafness, AI predicts how "natural" speech should sound. For visual impairment, AI tools such as Seeing AI and OrCam MyEye use images and video to predict what is nearby and narrate the environment (Granquist et al., 2021). AI may also enable smart wheelchairs for peo-

ple with mobility impairments (Giansanti and Pirrera, 2025), smart glasses to help people with autism spectrum disorder interpret emotional cues (Iannone and Giansanti, 2024), and brain–computer interfaces (e.g., Neuralink) that restore greater control to people with paralysis. Future work could examine the impact of such technologies on workplace outcomes.

We speculate that one reason the AI tool was very beneficial for DHH workers in our study is that the workers' disability is strongly affect them in one particular aspect of their job. That is, because many DHH workers have challenges with speech, they find it hard to communicate with customers, which is one part of their job. However, other disabilities in other jobs may have effects on multiple or many parts of the job, and such parts may not be amenable to AI. As such, we wish to be very careful about not overclaiming on the relevance of our results for disabled people in other jobs. Still, we believe there are many other jobs where a person's disability creates a localized bottleneck where AI might be useful.[32]

Relevance for other jobs. Our results are also specific to food delivery work, a job with short, frequent interactions with customers, decentralized work,[33] limited educational requirements, and heavy use of smartphones. Other jobs sharing these features include ride-hailing workers (e.g., Uber, Lyft, Didi); package delivery and last mile logistics (Amazon Flex, DoorDash Drive, UPS seasonal contractors); home services and on-demand repairs (TaskRabbit, Thumbtack, Handy); field sales or mechandising work (gig workers stocking shelves for Instacart, Shipt, or retail merchandising platforms); and customer-facing retail or hospitality roles where interactions are brief and standardized.

Public policy implications. There is significant public discussion about subsidizing AI to help the disadvantaged. However, in our particular setting, we estimate that it was *privately optimal* for the Platform to introduce the AI tool, with benefits much larger than costs. In such setting, it may be "optimal" for government to do little or nothing. There is a sizable number of DHH workers interested in working in gig jobs, especially in China, and the private benefits of simple AI tools for DHH workers is high.

Beyond gig workers with hearing impairments, firms may have incentives to implement similar technologies, without subsidies, when the disability is common and per-worker gains are high. That is, absent government intervention, AI may help reduce disability disparities on its own. For rarer disabilities, or where productivity improvements are smaller, public support may be warranted given the substantial social benefits of keeping disabled individuals

[32]For example, a blind / low-vision person might be fine with most aspects of knowledge work, but face a bottleneck in accessing visual information: s/he could benefit from an AI tool like Seeing AI. Likewise, someone with dyslexia may fare excellently in a marketing job, but face a bottleneck with regard to written expression (e.g., spelling errors), and Microsoft Copilot could be useful for this.

[33]In our context, food delivery workers in gig mode have no managers or colleagues. While Platform workers in team mode have managers, the work is still relatively decentralized compared to many jobs.

employed and reducing reliance on disability insurance.

While news stories often emphasize bad aspects of AI like job loss, AI tools like ours may help reduce inequality (Acemoglu et al., 2026).

The value of data. Separate from AI, our paper illustrates the value of granular administrative data in understanding the contributions of disabled workers. In analyzing the performance of DHH workers, one might naturally focus on their speed of work. While it is true that DHH workers are slower, they also work more hours and quit much less, consistent with having a worse outside option. We speculate that some organizations might "under-notice" this "loyalty benefit" of disabled workers. Big data can illuminate the ways in which disabled workers do or do not benefit organizations.

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.

_ , David Autor, and Simon Johnson, "Building Pro-Worker Artificial Intelligence," Working Paper 34854, National Bureau of Economic Research February 2026.

Adhvaryu, Achyuta, Martin Atela, Valentina Brailovskaya, Priyanka Dua, Jenny Susan John, Pratibha Joshi, and Rivandra Royono, "Platform Work: Evidence from Drivers in India, Indonesia, and Kenya," Technical Report, National Bureau of Economic Research 2026.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston, MA: Harvard Business Review Press, 2018.

Aizawa, Naoki, Corina Mommaerts, and Stephanie L. Rennane, "Firm Accommodation After Workplace Disability: Labor Market Impacts and Implications for Subsidy Design," *Econometrica*, 2025, *94* (2), 341–374.

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.

Arcot, Siva Venkatesh, "Comparative Analysis of AI-powered Outbound Dialer Campaigns vs. Legacy Outbound Dialer Campaigns (Contact Center)," *Asian Journal of Research in Computer Science*, Feb. 2025, *18* (3), 73–84.

Autor, David, "Polanyi's Paradox and the Shape of Employment Growth," in Lisa M. Lynch, ed., *Re-Evaluating Labor Market Dynamics*, Federal Reserve Bank of Kansas City, 2015, pp. 129–177.

_ , "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, Daniel Rock, and Chad Syverson, "The Productivity J-curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, 2021, *13* (1), 333–372.

_ , 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*, 2023, *139* (2), 891–936.

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.

Choi, Jung Ho and Chloe Xie, "Human+ AI in accounting: Early evidence from the field," 2025.

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.

Erickson, William, Camille Lee, and Sarah von Schrader, "Disability Statistics from the American Community Survey (ACS)," Cornell University Yang-Tan Institute (YTI), Ithaca, NY 2026. https://www.disabilitystatistics.org.

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.

Fruits, Eric and Kristian Stout, "AI, Productivity, and Labor Markets: A Review of the Empirical Evidence," Issue Brief ICLE Issue Brief, International Center for Law & Economics February 5 2026.

Gentry, Melissa, "Driving Inclusion: The Effect of Improved Transportation for People with Disabilities," *Unpublished Manuscript, Texas A&M*, 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.

Giansanti, Daniele and Antonia Pirrera, "Integrating AI and Assistive Technologies in Healthcare: Insights from a Narrative Review of Reviews," *Healthcare*, 2025, *13* (5), 556.

Granquist, Christina, Susan Y. Sun, Sandra R. Montezuma, Tu M. Tran, Rachel Gage, and Gordon E. Legge, "Evaluation and Comparison of Artificial Intelligence Vision Aids: Orcam MyEye 1 and Seeing AI," *Journal of Visual Impairment & Blindness*, 2021, *115* (4), 277–285.

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.

Iannone, Andrea and Daniele Giansanti, "Breaking Barriers—The Intersection of AI and Assistive Technology in Autism Care: A Narrative Review," *Healthcare*, 2024, *12* (2).

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.

Lane, Harlan, *The Mask of Benevolence: Disabling the Deaf Community*, New York: Alfred A. Knopf, 1992.

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 Conf. 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.

Merali, Ali, "Scaling Laws for Economic Productivity: Experimental Evidence in LLM-Assisted Translation," *arXiv preprint*, December 2024. Last revised Dec 7 2024.

Milligan, Kevin, "The Long-Run Growth of Disability Insurance in the United States," in "Social Security Programs and Retirement around the World," 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.

Olusanya, Bolajoko O., Adrian C. Davis, and Howard J. Hoffman, "Hearing Loss Grades and the International Classification of Functioning, Disability and Health," *Bulletin of the World Health Organization*, 2019, *97* (10), 725–728.

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.

Shah, Shariq, Hossein Ghomeshi, Edlira Vakaj, Emmett Cooper, and Shereen Fouad, "A Review of Natural Language Processing in Contact Centre Automation," *Pattern Analysis and Applications*, 2023, *26* (3), 823–846.

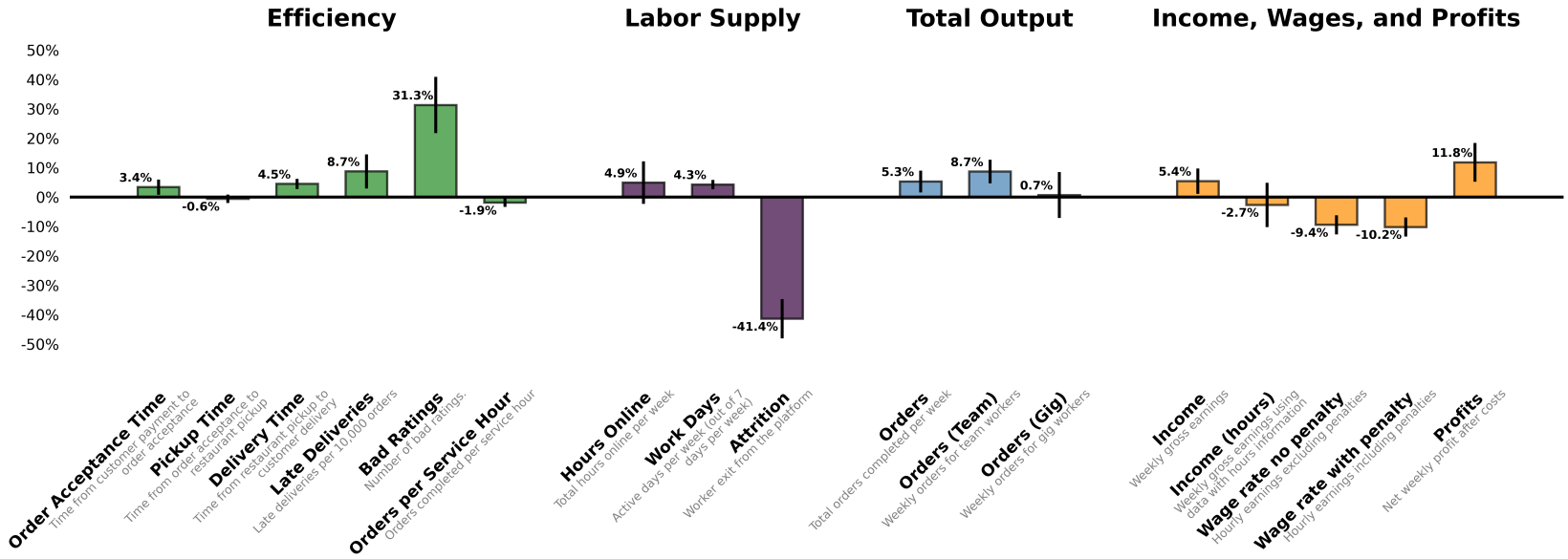
Shaw, Kathryn and Edward P. Lazear, "Tenure and Output," *Labour Economics*, 2008, *15* (4), 704–723.

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.

World Health Organization, "Deafness and hearing loss," https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss 2025. Accessed February 2026.

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.

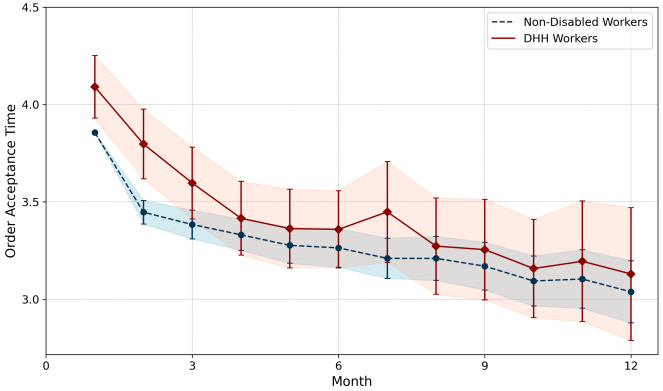
Labor supply notes: In the purple bars, *Hours Online* (total time the worker was online and available for orders, measured in hours per week), *Work Days* (number of days the worker was active in that week, out of 7), and *Attrition* reflects each worker's likelihood of exiting the Platform.

Efficiency notes: In the green bars, *Order Acceptance Time* is the dispatch time from order placement to a worker accepting it, *Pickup Time* is the travel time from accepting the order to arriving at the restaurant, and *Delivery Time* 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 Deliveries* is the number of late deliveries per 10,000 orders. *Bad Ratings* is the number of bad ratings in a week. *Orders per Service Hour* is orders completed per hour online.

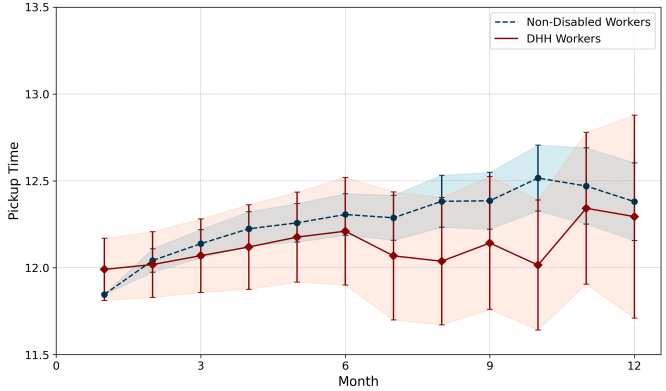
Total output notes: In the blue bars, *Orders* is the total number of delivery orders completed by a worker in a given week. We also show Orders results separately for gig and team workers.

Income, wages, and profits notes: In the orange bars, *Income* is a driver's weekly gross earnings. 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.

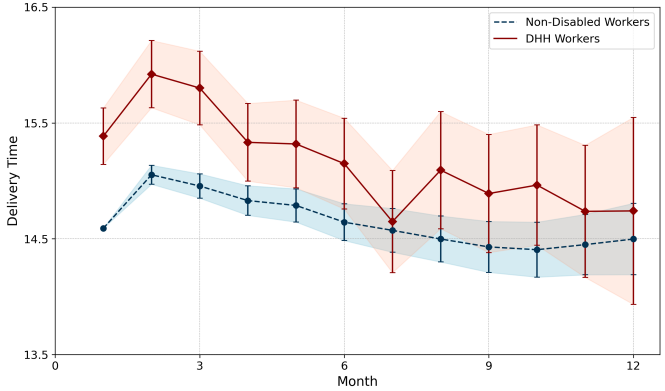
Figure 2: Pre-AI Tenure-Productivity Curves: Time Spent on Different Parts of the Order on Months of Tenure



(a) Order Acceptance Time



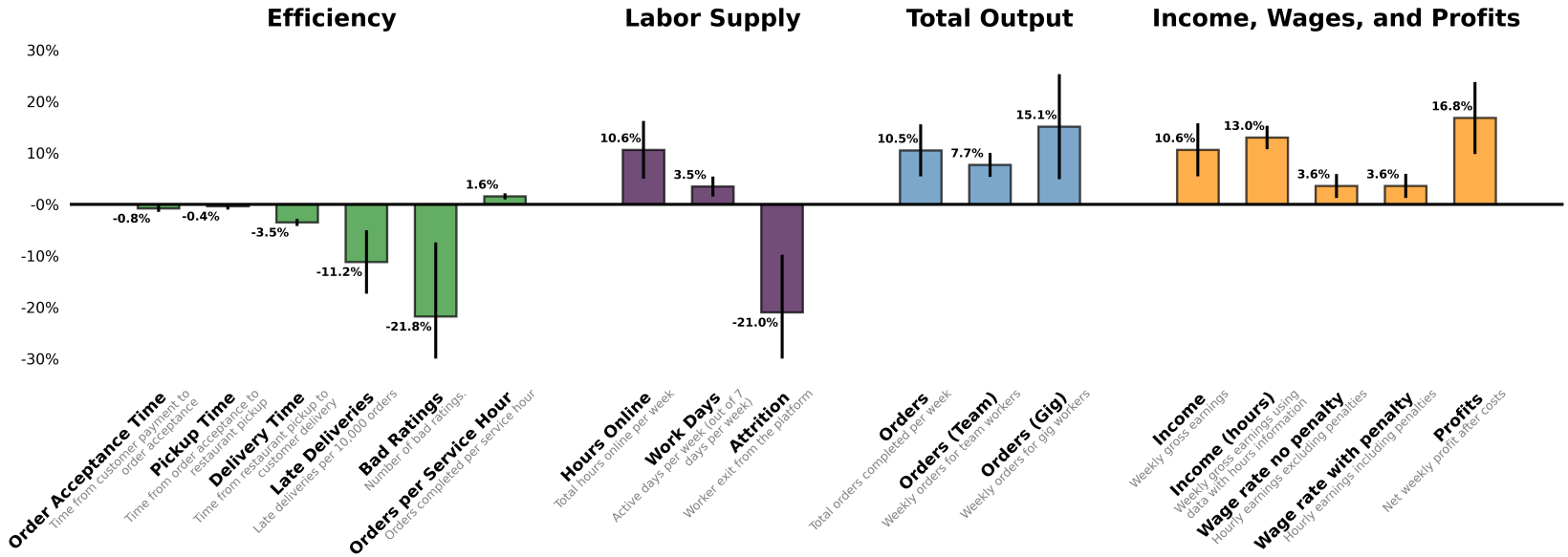
(b) Pickup Time



(c) Delivery Time

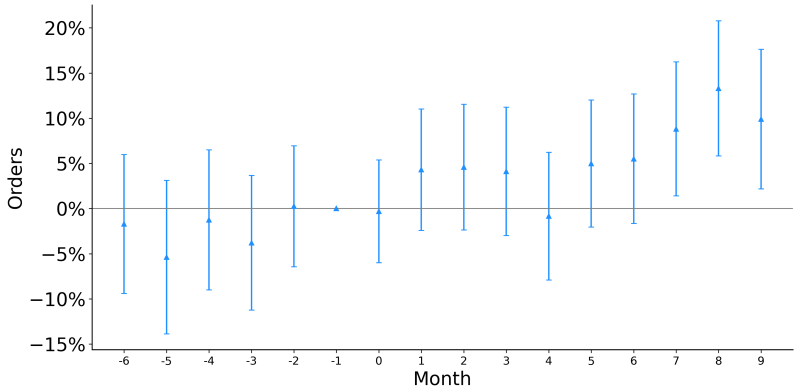
Notes: This figure examines how mean worker speed at different stages of the delivery process varies with worker tenure on the Platform. It plots time spent per order at different stages across each month of a worker's tenure, separately for DHH and non-disabled workers. The three time measures (order acceptance time, pickup time, and delivery time) are defined as in Figure 1. Each point on a curve represents the estimated average time at a given tenure month, obtained from the specification described in the main text. The red solid line corresponds to DHH workers, and the blue dashed line corresponds to non-disabled workers. The first month for non-disabled workers is normalized to zero; we then rescale all coefficients by adding back the average time for non disabled workers in their first month. Thus, at "Month 1," the figure shows the average per order time for workers in their first month of tenure. Error bars indicate 95% confidence intervals, with standard errors clustered at the worker level.

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

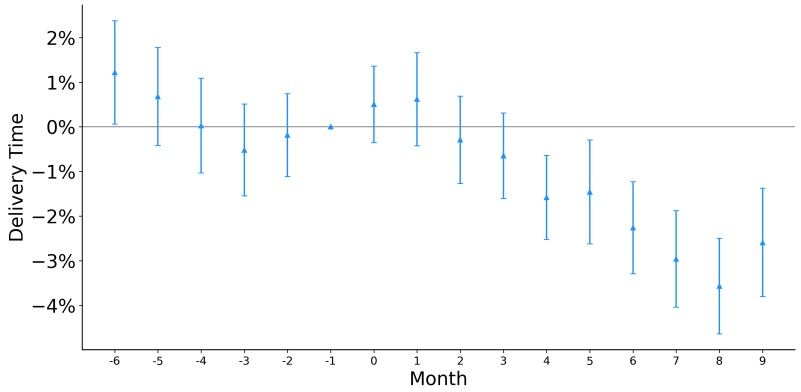


Notes: This figure shows DiD estimates of the impact of the AI-based Intelligent Outbound Call tool on DHH workers. Each bar reports the percentage impact of the AI tool on gaps between DHH and non-disabled workers. Positive values indicate improvements for DHH workers relative to non-disabled workers (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.

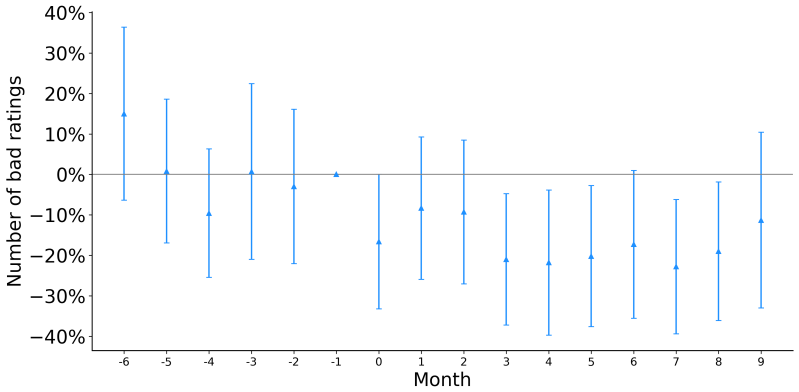
Figure 4: Event Study Analysis: AI Tool Effects on Productivity



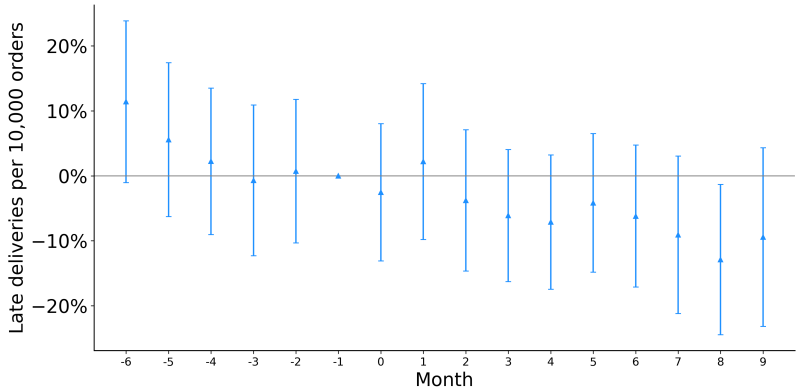
(a) Number of Orders Completed



(b) Delivery Time



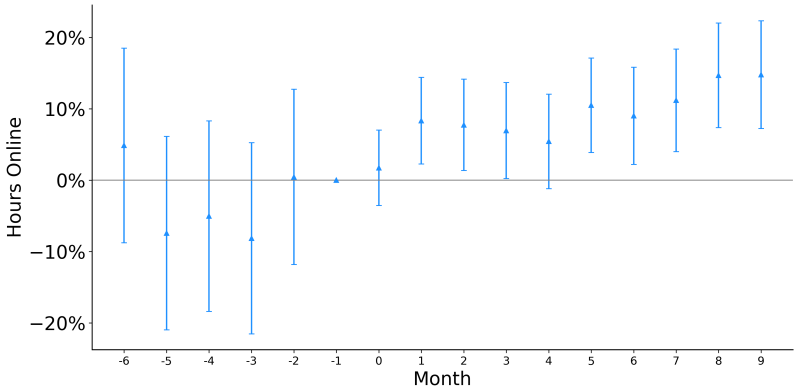
(c) Bad Ratings



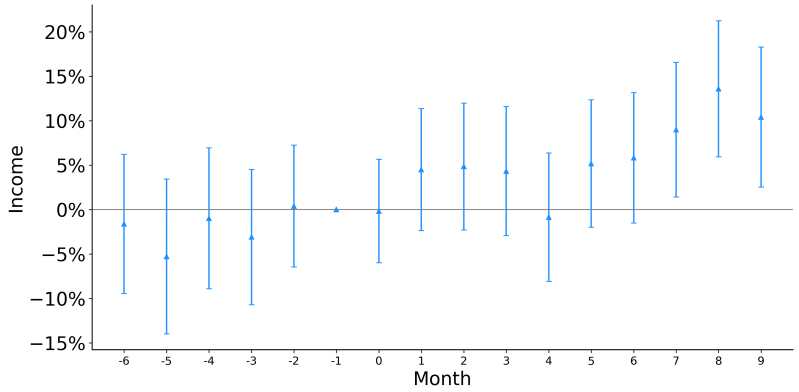
(d) Late Deliveries

Notes: This figure presents an event study analysis of the introduction of the AI-based Intelligent Outbound Call (IOC) system. The x-axis shows event time in months relative to the intervention, with 0 denoting the first month in which the tool is in effect. The underlying data are weekly; months are defined as 4-week periods, anchored on the exact week of implementation. Dots show the event study coefficients, and whiskers show the 95% confidence intervals accounting for clustering by driver. The y-axis shows the estimated difference between DHH and non-disabled workers. Panel (a) shows effects on weekly orders. Panel (b) shows effects on delivery time. Panel (c) shows effects on the number of bad ratings. Panel (d) shows effects on late deliveries.

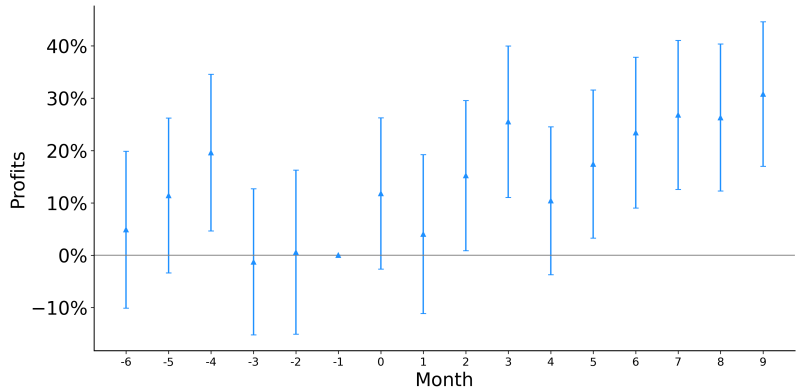
Figure 5: Event Study Analysis: AI Tool Effects on Hours, Income, and Profits



(a) Hours Worked



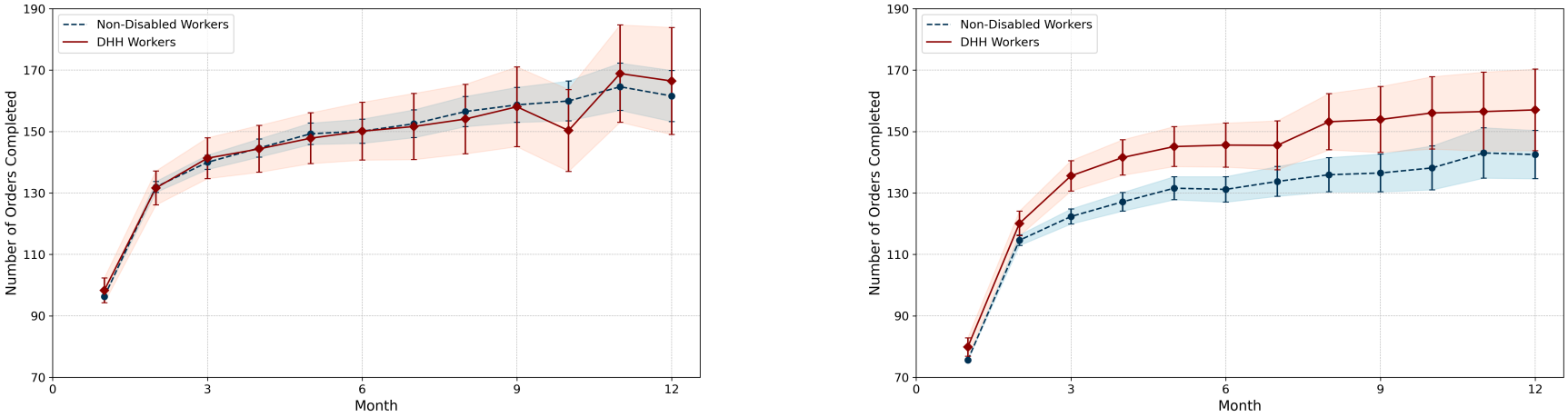
(b) Income



(c) Profits

Notes: This figure is similar to Figure 4, but examines the effects of the AI tool on DHH workers' hours worked, income, and platform profits. All these variables are defined in the main text and also described in Figure 1. Hours worked is based on worker-weeks with hours data. Income and profits are based on all the data and do not restrict to worker-weeks with hours data. Dots show the event-study coefficients, and whiskers show the 95% confidence intervals, accounting for clustering by worker.

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

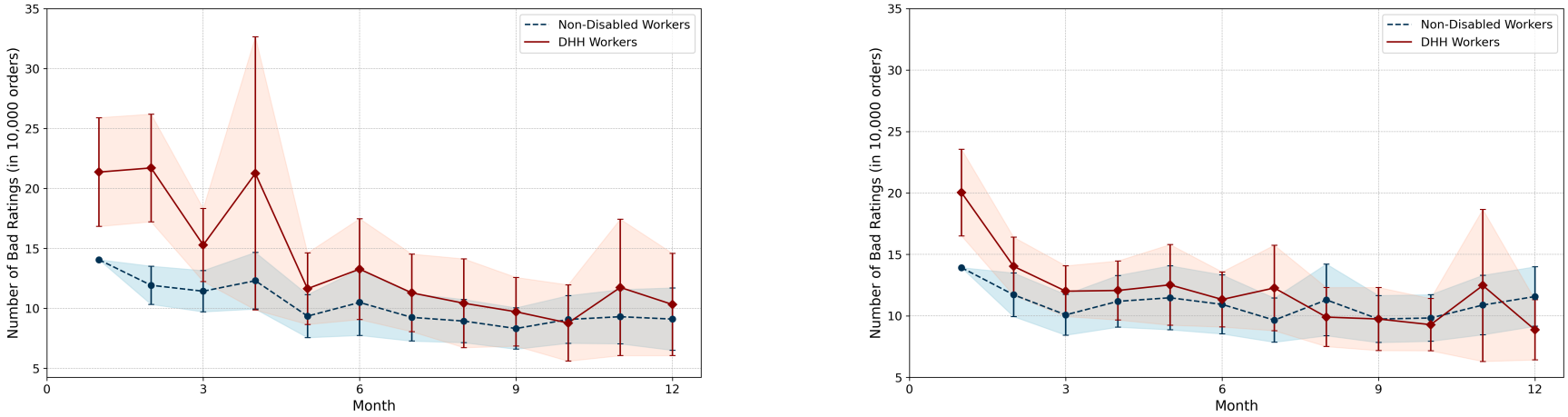


(a) Pre-AI tool

(b) Post-AI tool

Notes: This figure examines how the productivity trajectory of new workers over their first year differs before and after the introduction of the AI tool. The figure plots the average number of orders completed as a function of months of tenure. Error bars show 95% confidence intervals accounting for clustering by worker. The red solid line corresponds to DHH 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 split by whether workers had access to the AI tool from the start of their tenure. The AI tool was implemented in mid 2022 at time t_0. Panel (a), *Pre-AI tool*, displays tenure profiles for workers who entered the Platform prior to the AI tool, and we include their observations until the time of AI implementation. For example, if someone joins the Platform at time $t_0 - 2$, we include two months of tenure for this worker, corresponding to months $t_0 - 2$ and $t_0 - 1$. The Panel (a) sample includes 191,139 worker-month observations from 17,732 workers, of whom 1,624 are DHH and 16,108 are non-disabled. Panel (b), *Post-AI tool*, includes workers who enter the Platform after the AI tool is introduced. The Panel (b) sample includes 186,760 worker-month observations from 25,873 workers, of whom 2,660 are DHH and 23,213 are non-disabled.

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

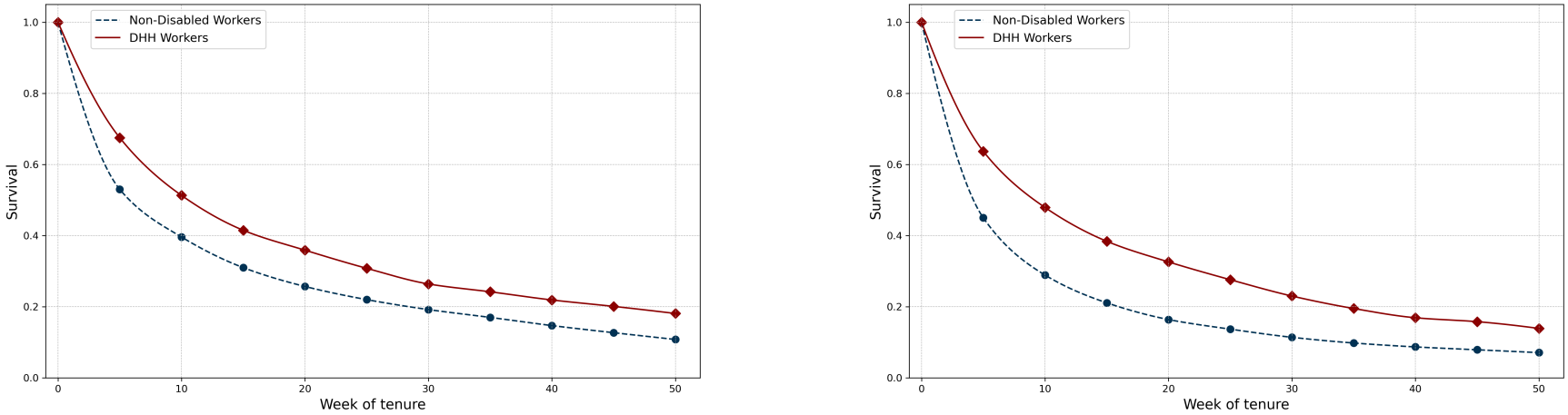


(a) Pre-AI tool

(b) Post-AI tool

Notes: This figure is similar to Figure 6, but the outcome is bad customer reviews, measured in terms of bad customer reviews per 10,000 orders. Error bars show 95 percent confidence intervals accounting for clustering by worker. The samples in each panel are the same as in Figure 6.

Figure 8: Survival Curves: Before and After the AI Tool

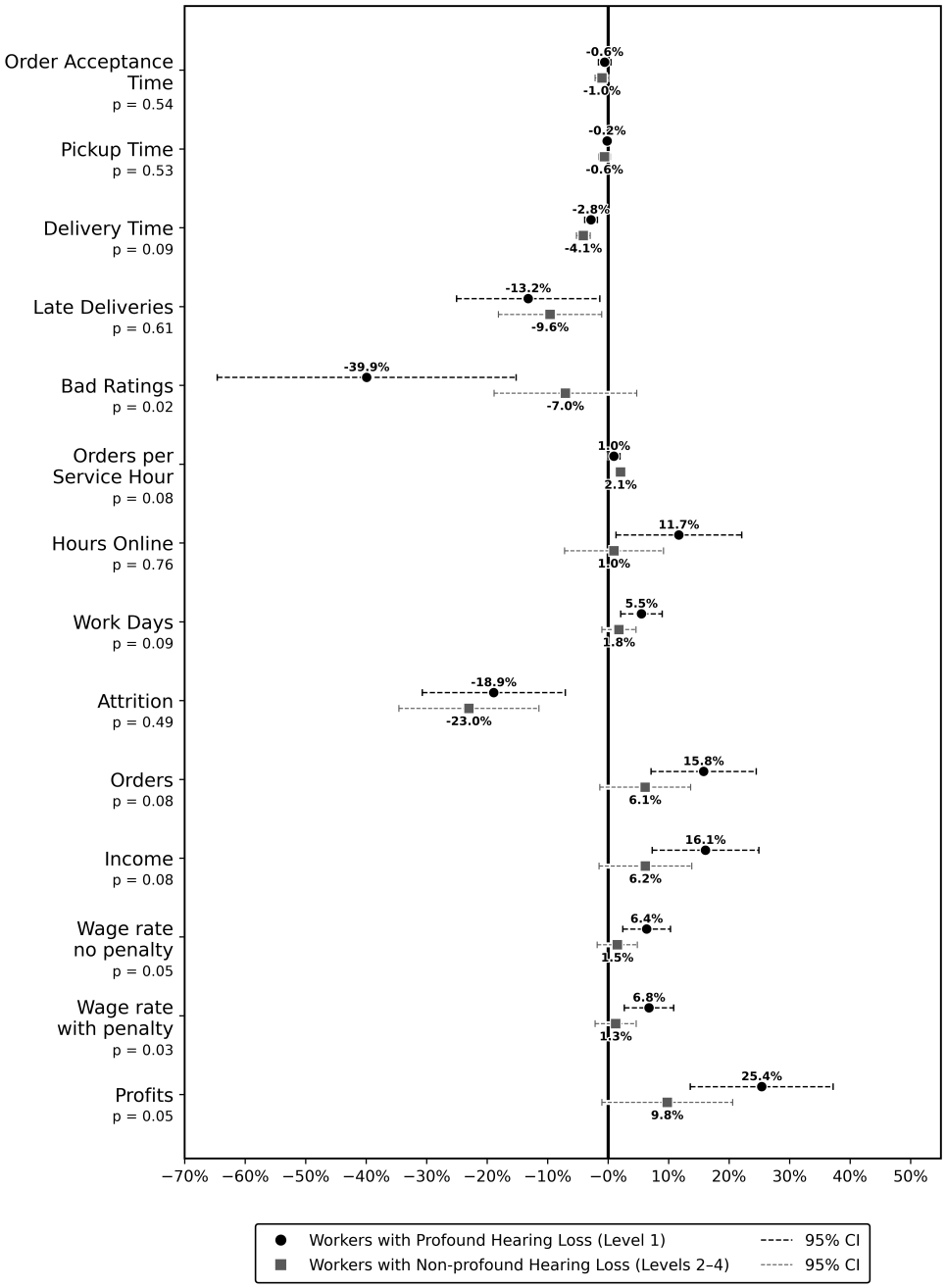


(a) Pre-AI tool

(b) Post-AI tool

Notes: This figure is similar to Figure 6, but shows survival curves as the outcome. 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 samples in each panel are the same as in Figure 6.

Figure 9: Heterogeneous Treatment Effects of the AI Tool by Disability Severity



Notes: This figure examines how impacts of the AI tool vary according to disability severity. Each dot-and-whisker plot is calculated using a regression where DHH workers of a given severity are compared to non-disabled workers. Each dot or square represents the treatment effect of the AI tool on DHH workers of a given severity. The whiskers show 95% confidence intervals accounting for clustering by driver. The p-values shown are from tests of whether there is a differential effect by disability severity. Each p-value is calculated from one regression with two key interaction terms, Profound $\text{DHH}_i \times Post_t$ and Non-profound $\text{DHH}_i \times Post_t$, testing whether the two interactions are different, where Profound DHH_i and Non-profound DHH_i are dummies for having profound hearing and non-profound hearing loss, respectively. As described in Section 1, severity follows the official Chinese four-level classification (GB/T 26341-2010), where Level 1 is profound hearing loss, corresponding to more than 90 decibels of hearing loss, and Level 4 is the least severe (41-60 decibels of hearing loss). Severity level is based on DHH drivers' government certificate of disability.

Table 1: Mean Characteristics of DHH and Non-disabled Workers

	DHH Workers (1)	Non-Disabled Workers (2)	Full Sample (3)	P-Value of Difference (4)
Panel A: Worker Demographics				
Age	32.8	30.3	30.5	0.00
Male	92.6%	90.8%	91.0%	0.00
Panel B: Work Model				
Team worker	35.4%	36.4%	36.3%	0.18
Gig worker	64.6%	63.6%	63.7%	0.18
Panel C: Disability Severity for DHH Workers				
Level 1: Profound hearing loss	53.6%			
Level 2: Severe hearing loss	16.5%			
Level 3: Moderate-Severe hearing loss	14.2%			
Level 4: Moderate hearing loss	15.7%			
Panel D: Distance & Geography Variables				
Driver is in a big city	36.6%	32.3%	32.7%	0.00
Driver is in Shanghai	7.27%	5.58%	5.75%	0.00
Driver is in Beijing	3.60%	3.36%	3.38%	0.40
Driver is in Eastern China	64.2%	58.2%	58.8%	0.00
Avg. distance bwn. restaurant and customer, in km	2.24	2.16	2.17	0.00
Unique restaurants served, per week	73	65	66	0.00
Number of workers	4,284	39,321	43,605	
Number of worker-weeks	70,301	409,769	480,070	

Notes: This table provides summary statistics on food-delivery workers on the Platform, both deaf or hard of hearing (DHH) and non-disabled. We use data from the whole sample period of 2021-2023. Column (1) shows DHH worker averages; (2) shows non-disabled worker averages; (3) is the full sample average. Column (4) reports the p-value for the difference in means between DHH and non-disabled worker. The unit of observation is a worker for demographics. In Panel A, age is age upon joining the Platform. In Panel B, Team worker and Gig worker are dummies for a driver's work model, also called their registration type. Panel C reports the distribution of disability severity among the DHH workers in column 1. Severity follows the official four-level classification (GB/T 26341-2010). Level 1 is profound hearing loss, meaning loss of greater than 90 decibels; Level 2 is severe hearing loss of 81-90 decibels; Level 3 is moderately severe hearing loss of 61-80 decibels; and Level 4 is moderate hearing loss of 41-60 decibels. Level 4 is the lowest level of severity to have a certificate of disability in China. Severity is based on government medical certificates of disability, which are provided by drivers to the Platform. Big cities cover the top 10 cities in China by population: Chongqing, Shanghai, Beijing, Chengdu, Guangzhou, Shenzhen, Wuhan, Tianjin, Xi'an, and Suzhou. Eastern China provinces include: Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan.

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					
Order Acceptance Time: minutes from customer payment to order acceptance	3.37	2.48	9.20	464,717	42,281
Pickup Time: minutes from order acceptance to restaurant pickup	11.97	11.47	6.86	464,850	42,293
Delivery Time: minutes from restaurant pickup to customer delivery	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 Week	0.09	0.00	0.36	480,070	43,605
Week where worker gets at least one Bad Rating, 0 or 1	0.07	0.00	0.26	480,070	43,605
Good Ratings per Week	9.32	6.00	12.30	480,070	43,605
Week where worker gets at least one Good Rating, 0 or 1	0.81	1	0.38	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. We use data from the whole sample period of 2021-2023. 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.

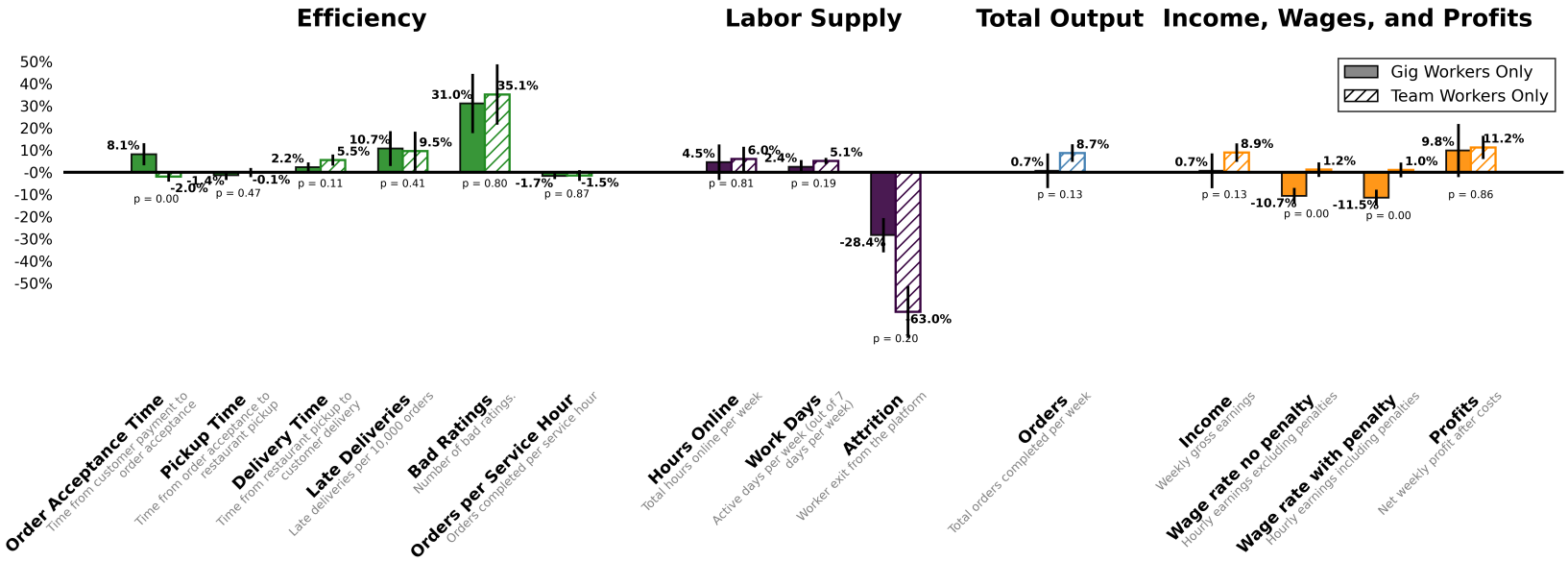
Web Appendix for "Empowering Inclusive Work"

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

The Web Appendix is organized as follows. Appendix A provides additional figures and tables. Appendix C analyzes how the AI tool affects new signups (i.e., entry) of DHH workers onto the Platform. Appendix D describe the qualitative data collected on DHH workers on the Platform. Appendix E provides further information on hearing loss levels in China. Appendix F discusses issues related to potential nondisclosure of DHH status by DHH workers on the Platform. Appendix G provides further details and visuals on the AI tool.

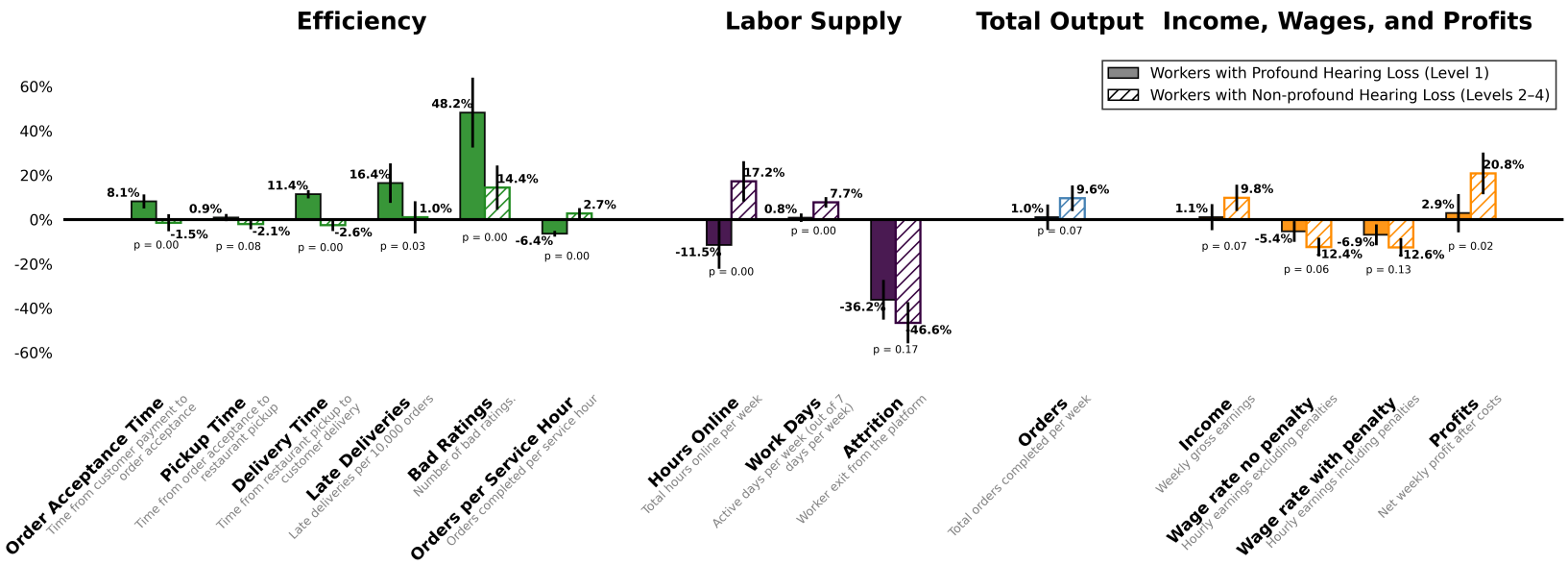
A Appendix Figures and Tables

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



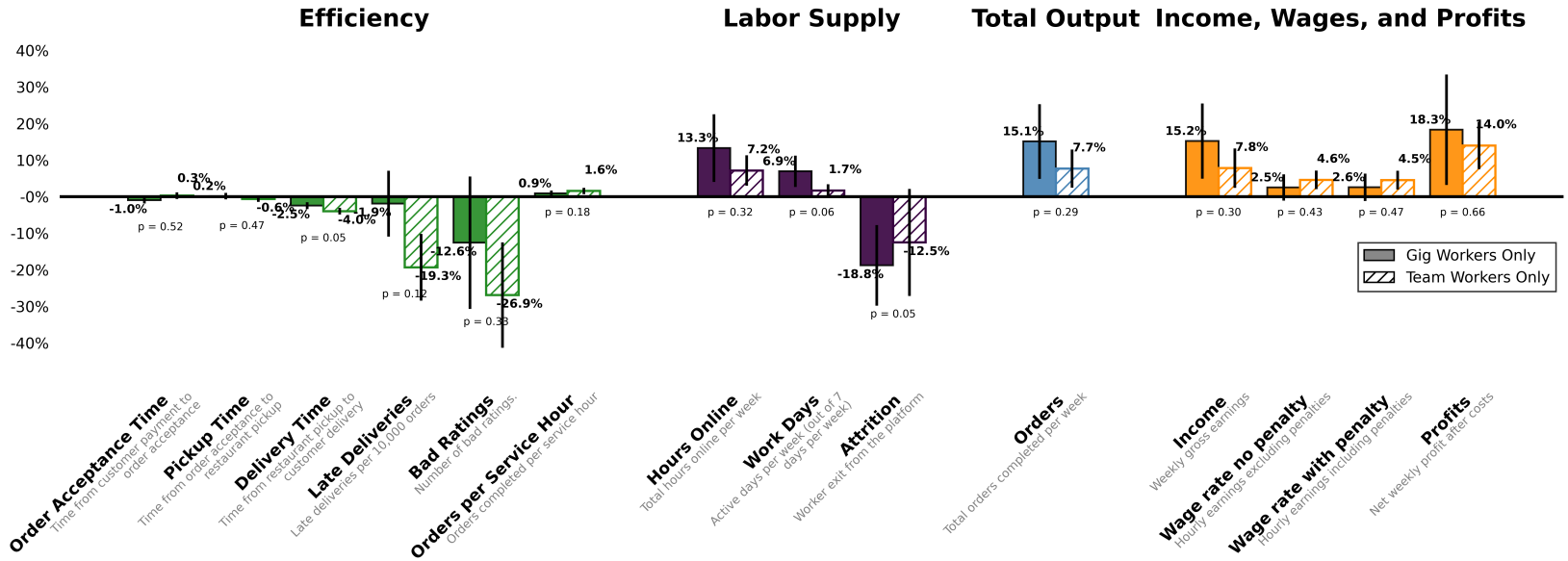
Notes: This figure reproduces the analysis in Figure 1, but splits the sample by gig workers and team workers. We plot pre-AI differences for DHH gig workers vs. non-disabled gig workers, as well as pre-AI differences for DHH team workers vs. non-disabled team workers. Each bar is estimated from a separate regression in the corresponding subsample (gig or team workers). P-values are based on pooled regressions with interaction terms that test equality of the underlying OLS coefficients (in levels) across groups. Because the plotted coefficients are scaled by group-specific pre-AI means, while the hypothesis tests are conducted on the underlying level differences, the p-values need not correspond exactly to visual differences in the bars. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept. The p-value for the attrition result corresponds to a test based on the raw OLS percentage-point differences, rather than the plotted scaled coefficients.

Figure A.2: Pre-AI Differences Between DHH and Non-Disabled Workers: Splitting by Severity of Disability



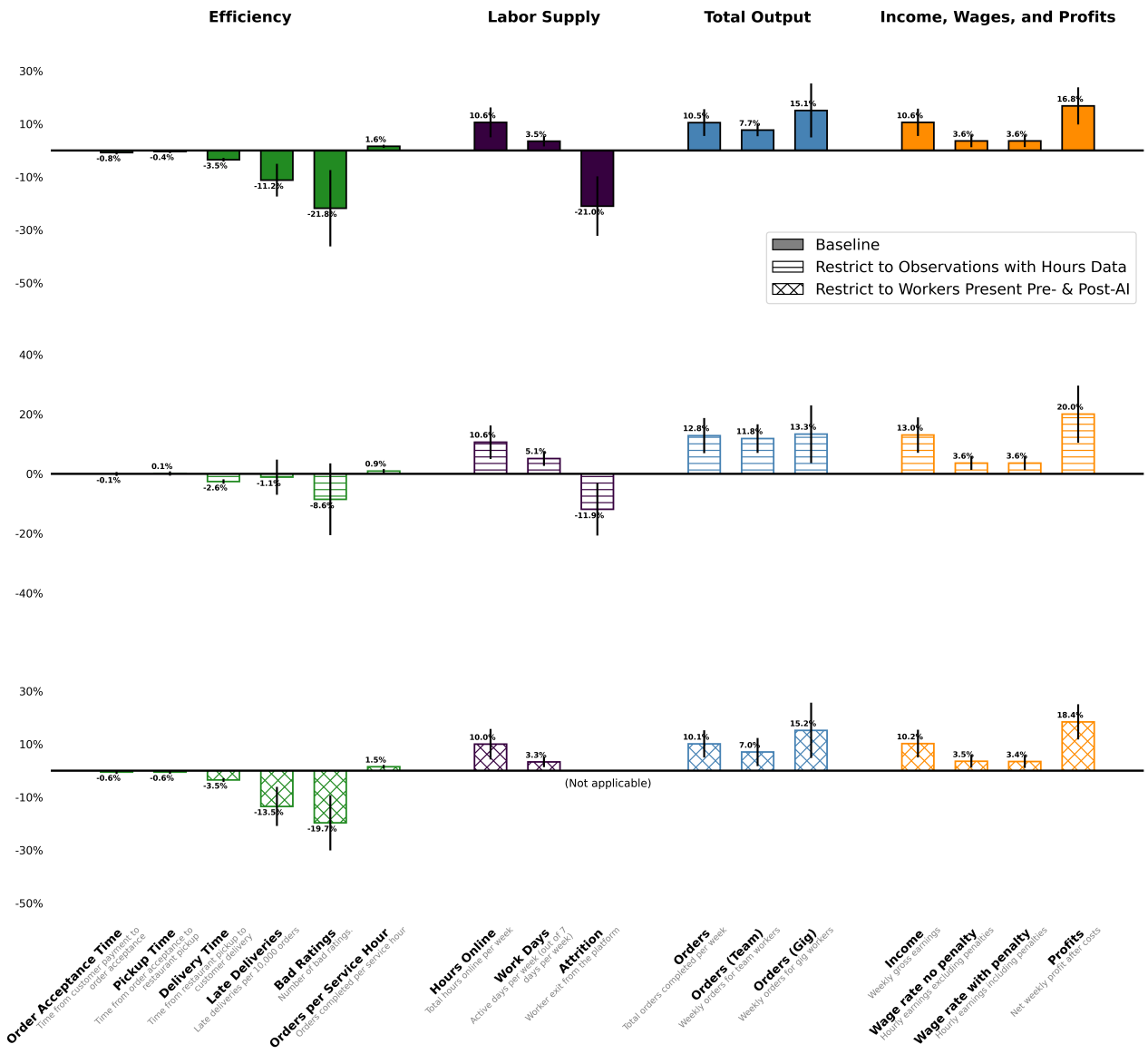
Notes: This figure reproduces the analysis in Figure 1, but splits DHH differences by severity of disability. As discussed in Section 1 of the main text, profound hearing loss (Level 1 in the Chinese system) is loss of more than 90 decibels. Non-profound hearing loss is loss of 41-90 decibels.

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



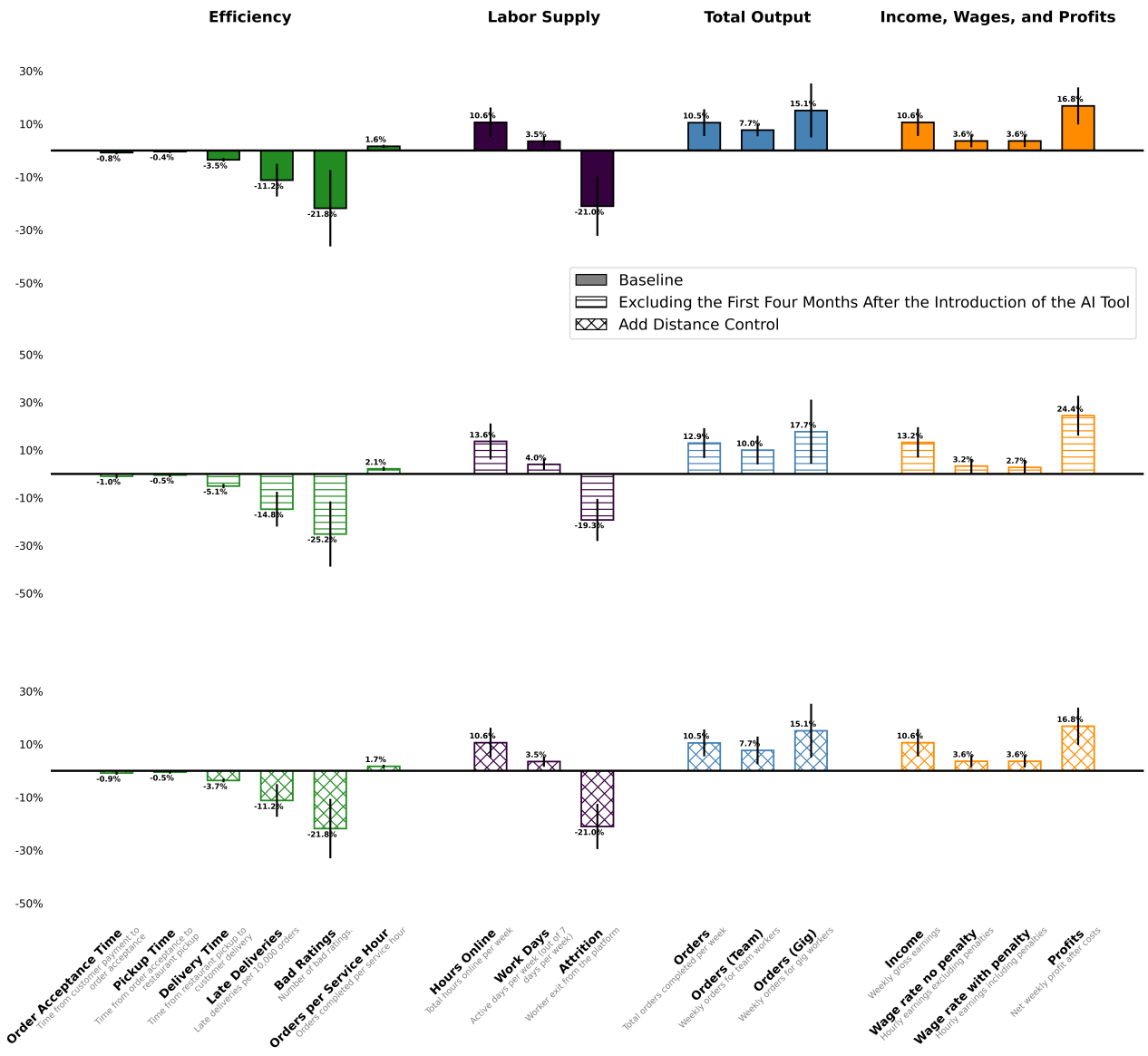
Notes: This figure reproduces the analysis in Figure 3, focusing exclusively on gig or team workers. As discussed in the main text, gig workers operate individually, exercise full discretion over their working hours, and decide which orders to accept.

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



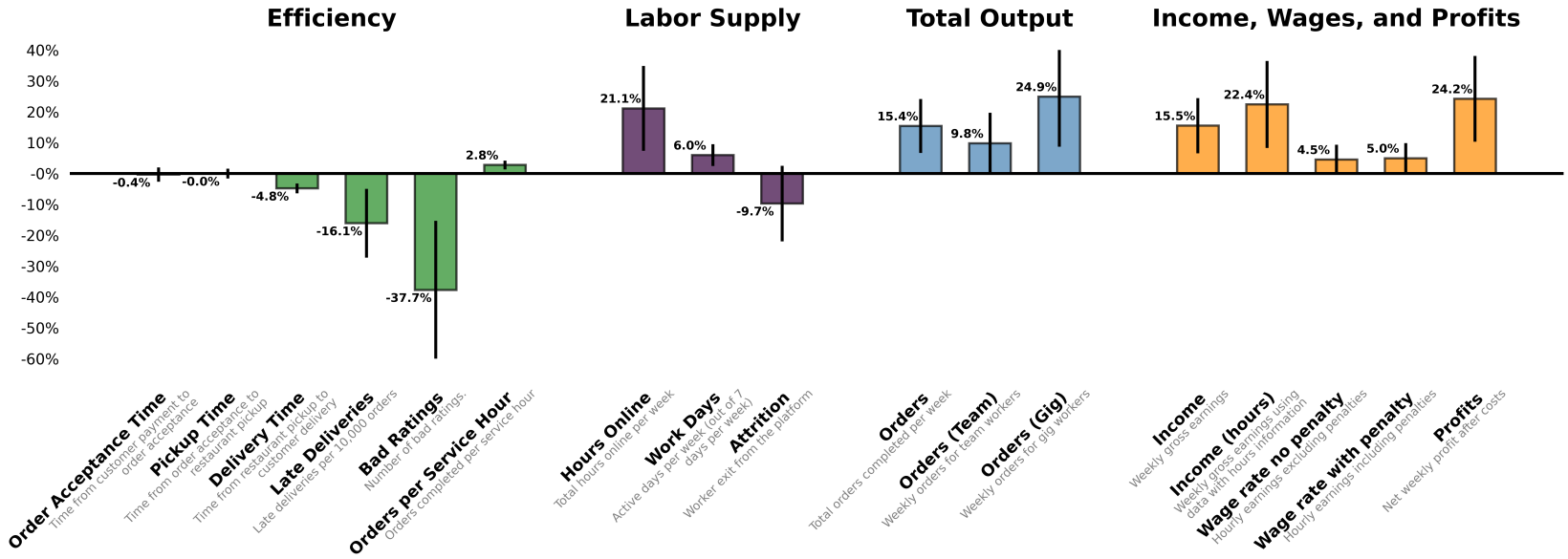
Notes: This figure replicates the analysis in Figure 3. The top panel reports the baseline estimates using the full sample. The middle panel restricts the sample to observations with hours data, addressing concerns related to measurement of labor supply. The bottom panel restricts the sample to drivers 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



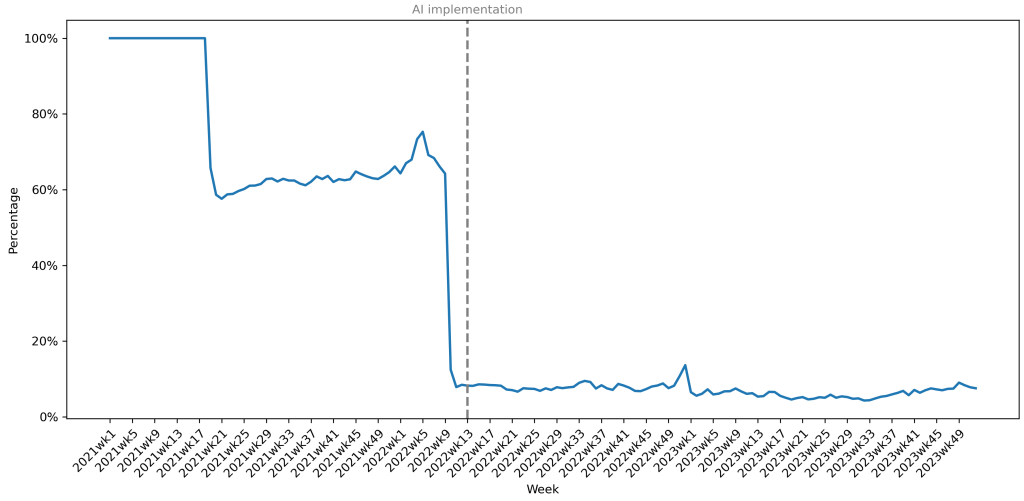
Notes: The top panel reports the baseline estimates. The middle panel excludes the first four months after the introduction of the AI tool for potential learning and onboarding effects. The bottom panel adds controls for task difficulty, proxied by order distance. All specifications follow the same difference-in-differences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

Figure A.6: Robustness of Figure 3 to Excluding COVID Lockdown Periods

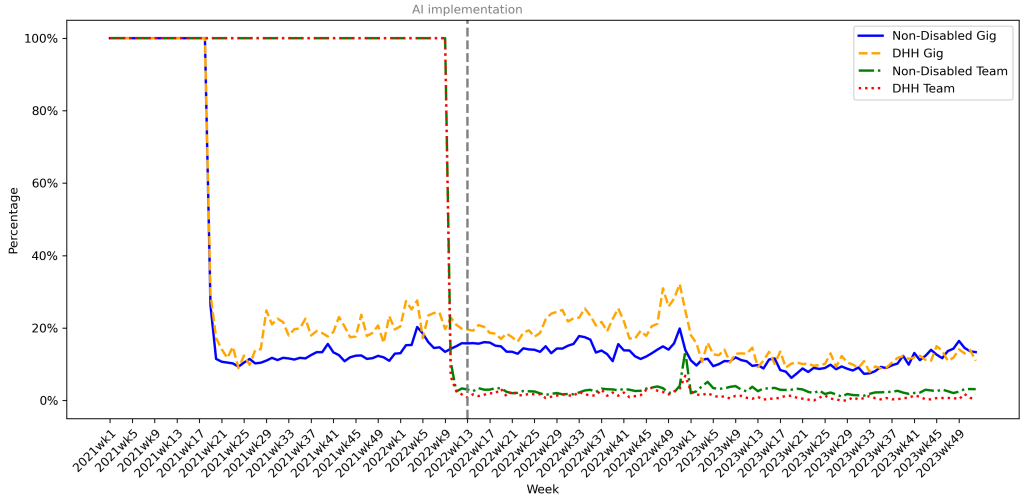


Notes: All specifications follow the same difference-in-differences design as in Figure 3. This figure excludes all worker-week observations that fall in province-weeks affected by COVID outbreaks. Under China's 'Dynamic Zero-COVID' policy, restrictions were implemented through localized measures rather than a single synchronized nationwide lockdown, so the timing and stringency of restrictions varied across regions. We identify affected province-weeks using official government announcements and verify the timing with contemporaneous news reports.

Figure A.7: The Share of Worker-Weeks with Missing Hours Data Over Time



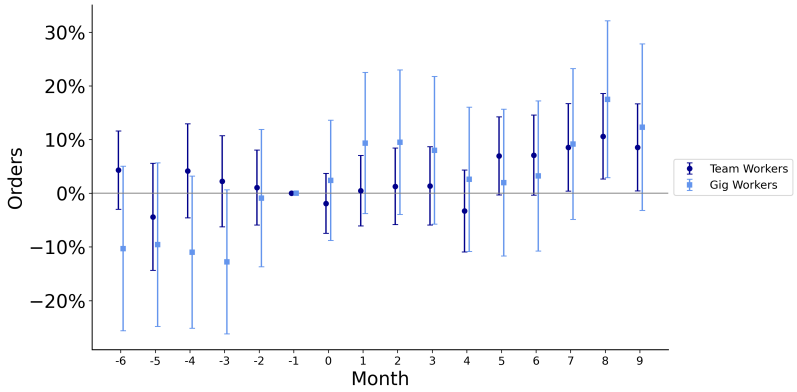
(a) Share of worker-weeks with missing hours over time



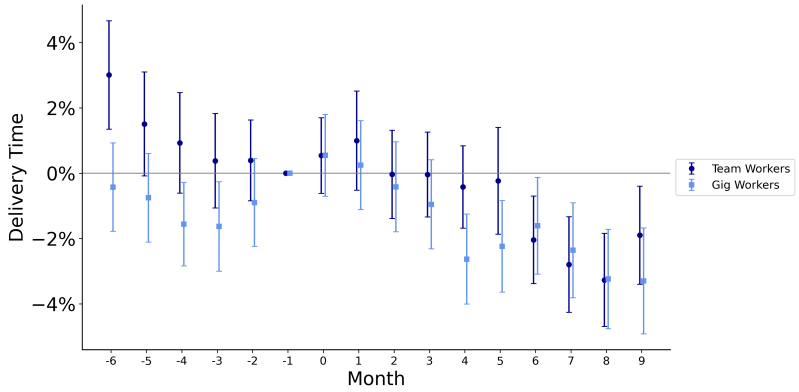
(b) Share of worker-weeks with missing hours, split by work model and DHH status

Notes: This figure shows the share of worker-weeks with missing hours data over time. The dotted vertical line shows when the AI tool is implemented. 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 become available for gig workers in mid 2021, and then for team workers in early 2022.

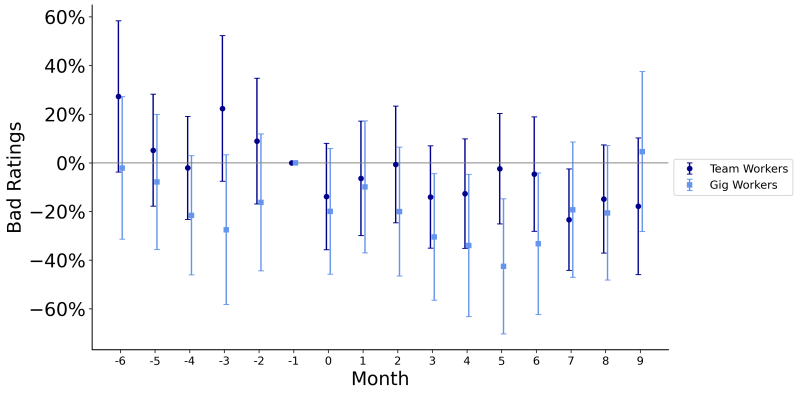
Figure A.8: Event Study Analysis on Efficiency: Splitting by Gig vs. Team Workers



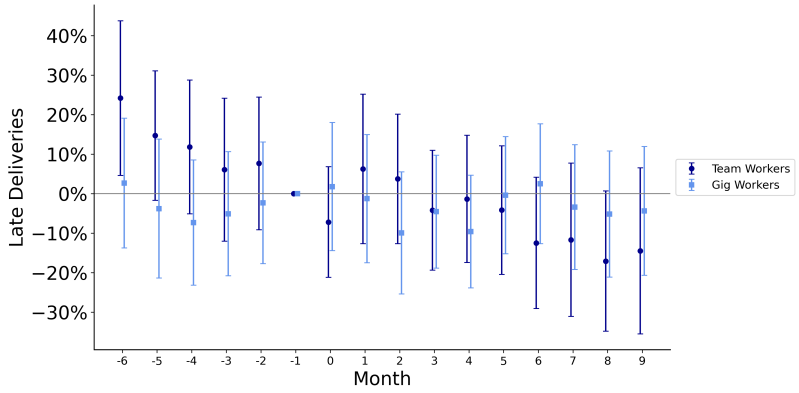
(a) Number of Orders Completed



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



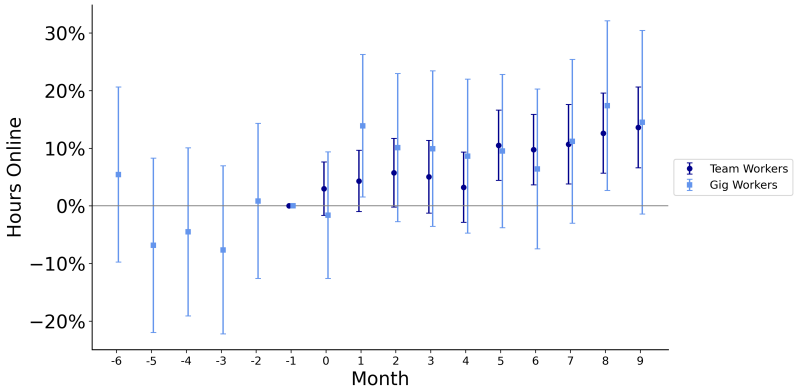
(c) Bad Ratings



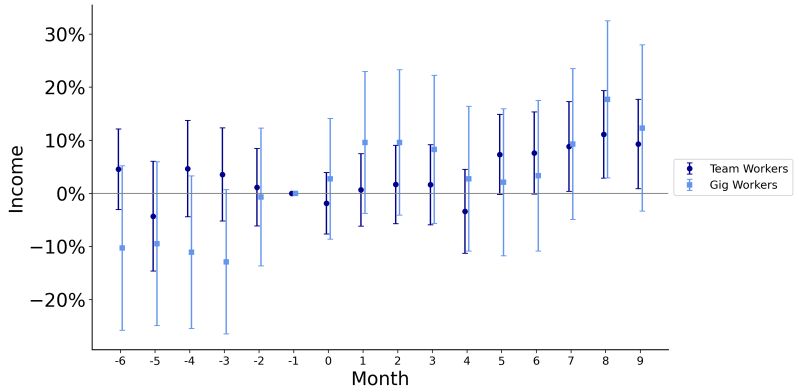
(d) Number of Delayed Orders

Notes: This figure is similar to Figure 4 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.

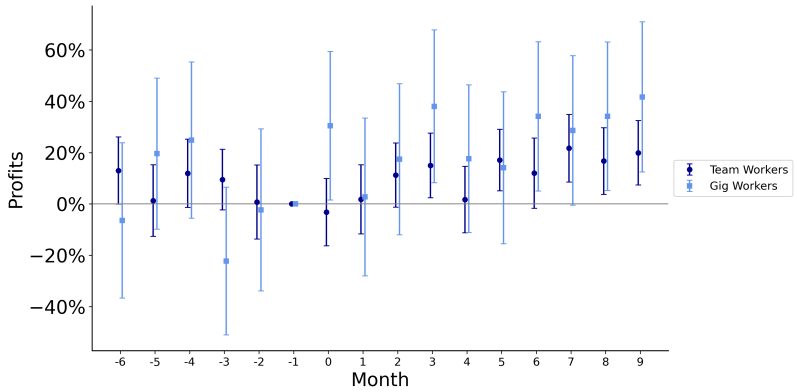
Figure A.9: Event Study Analysis on Hours, Income, and Profits: Splitting by Gig vs. Team



(a) Hours Worked



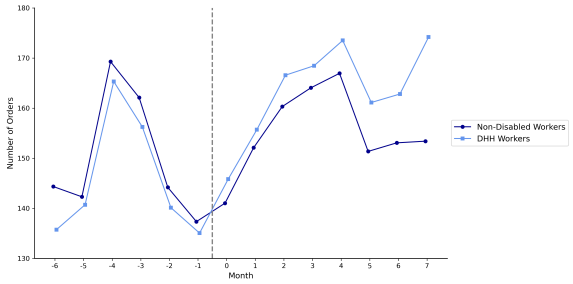
(b) Income



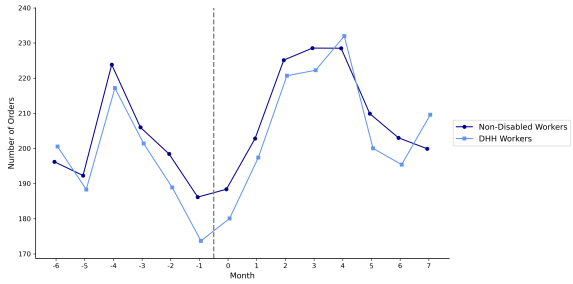
(c) Profit

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.

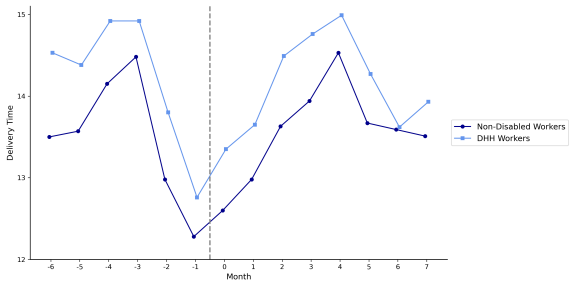
Figure A.10: Comparison of Raw Means Between DHH and Non-Disabled Workers



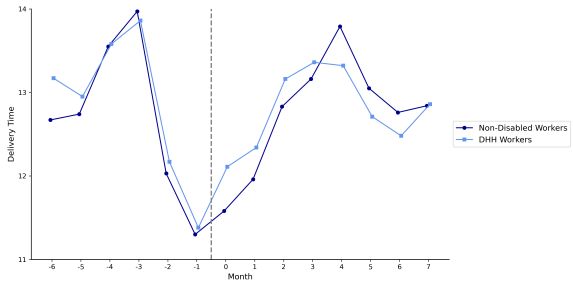
(a) Orders Completed
(All Workers)



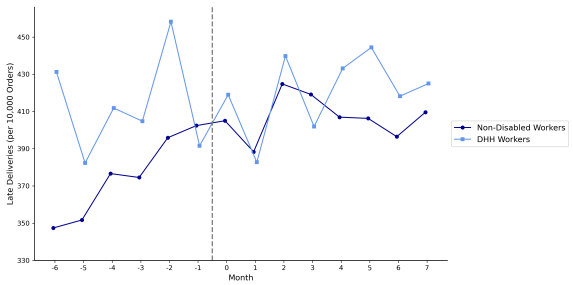
(b) Orders Completed
(Balanced Panel)



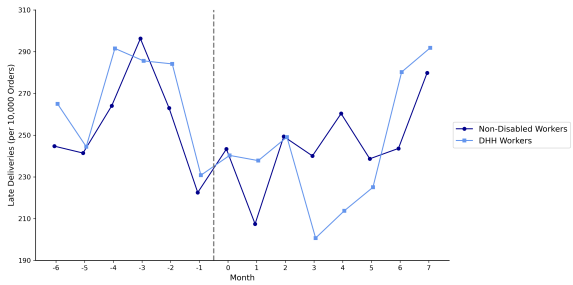
(c) Delivery Time
(All Workers)



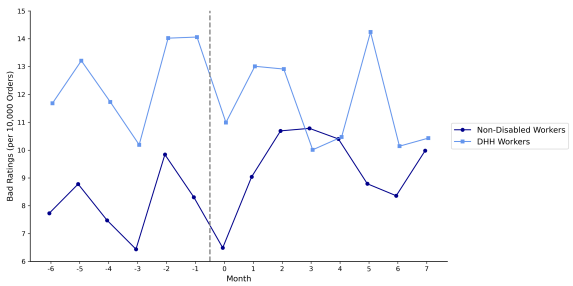
(d) Delivery Time
(Balanced Panel)



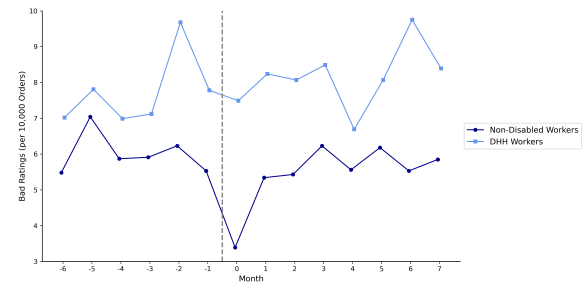
(e) Late Delivery
(All Workers)



(f) Late Deliveries
(Balanced Panel)



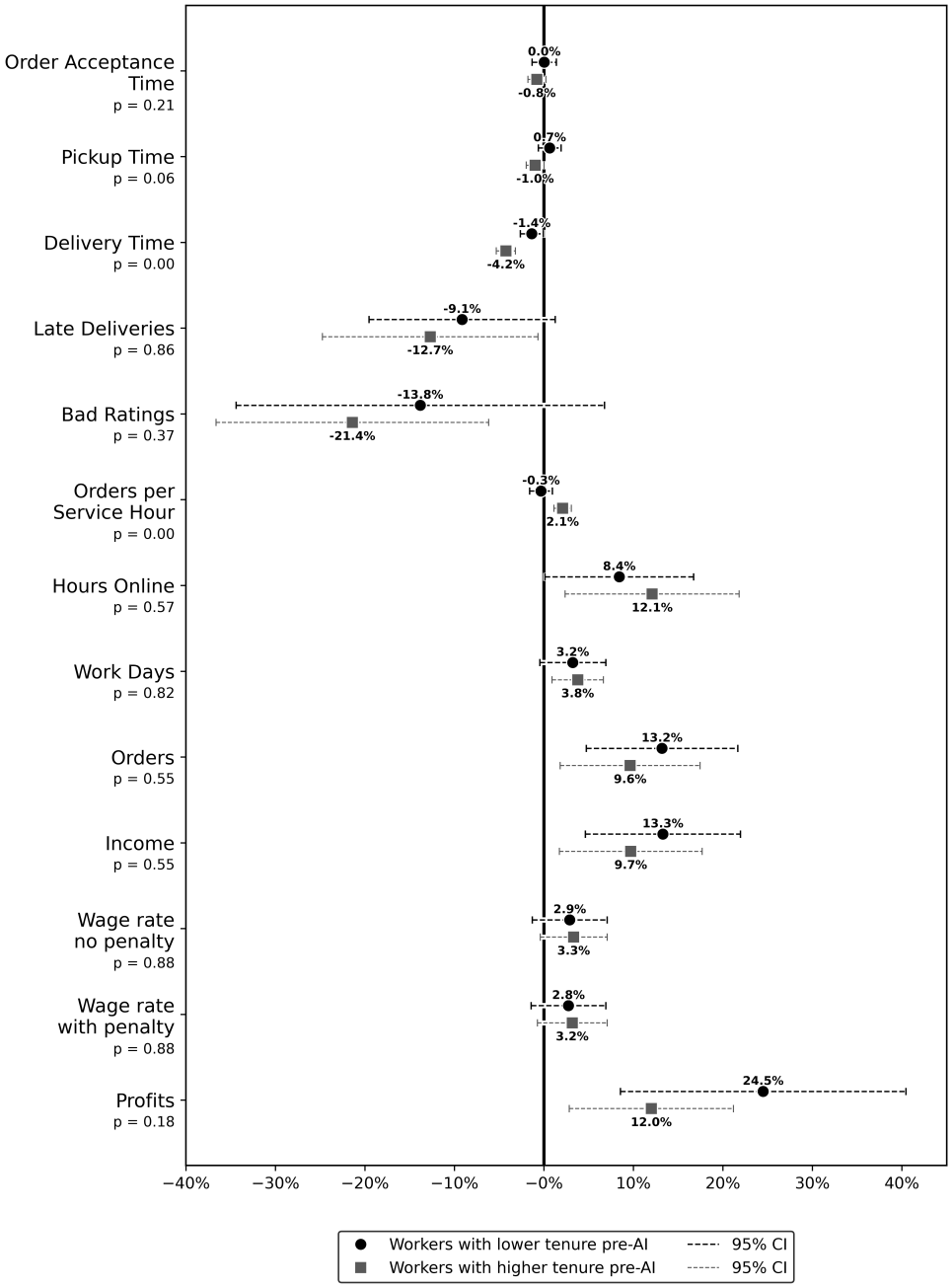
(g) Bad Ratings
(All Workers)



(h) Bad Rating
(Balanced Panel)

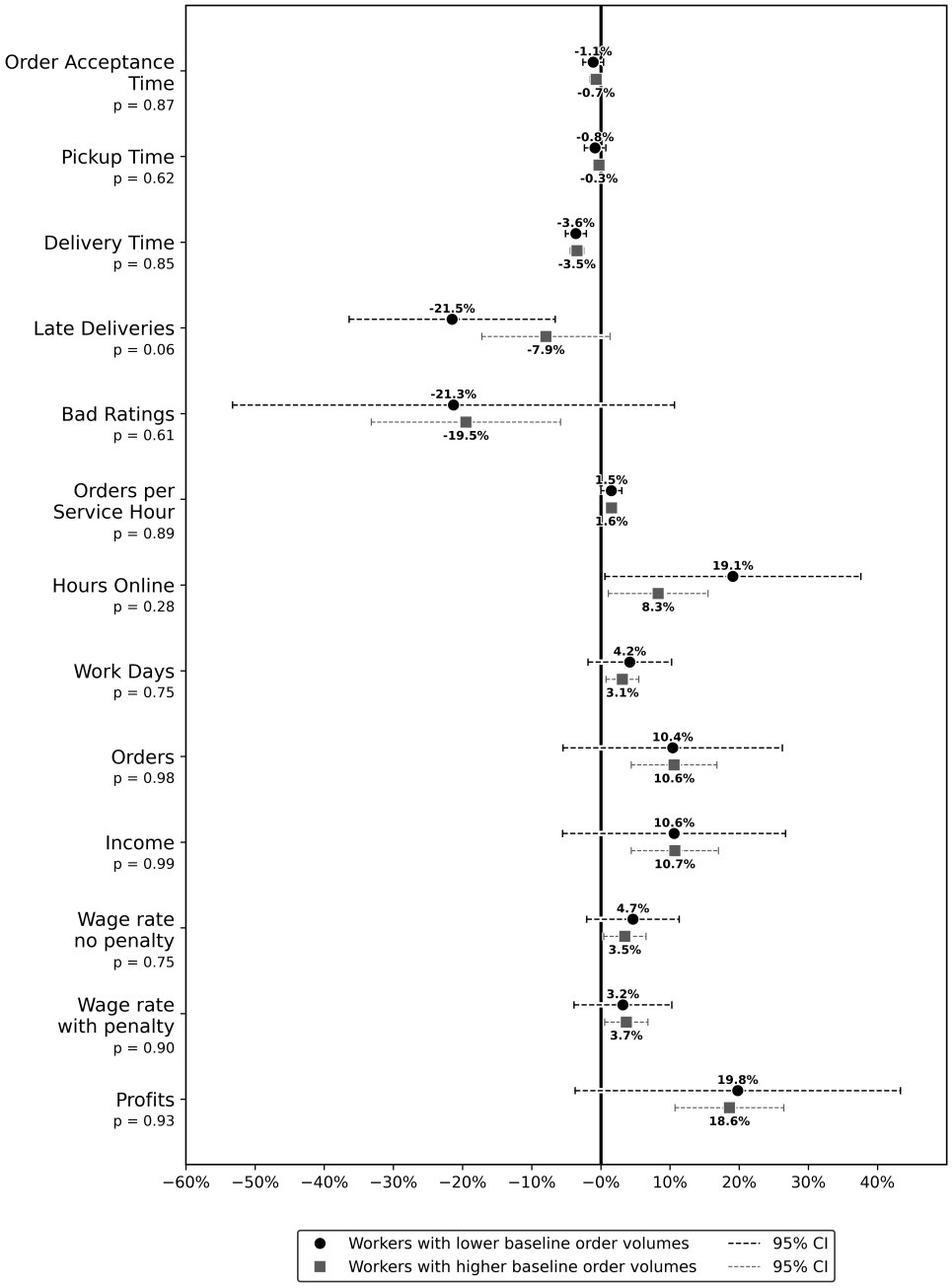
Notes: This figure compares raw means of outcomes for DHH workers (light blue) and non-disabled workers (dark blue). The left panels show results for all workers. There are 43,605 workers in total, including 4,284 DHH workers and 39,321 non-disabled workers. The right panels show results for workers 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 779 workers in total, with 107 DHH workers and 672 non-disabled workers.

Figure A.11: Heterogeneity in Effects of AI Tool by Pre-AI Tenure



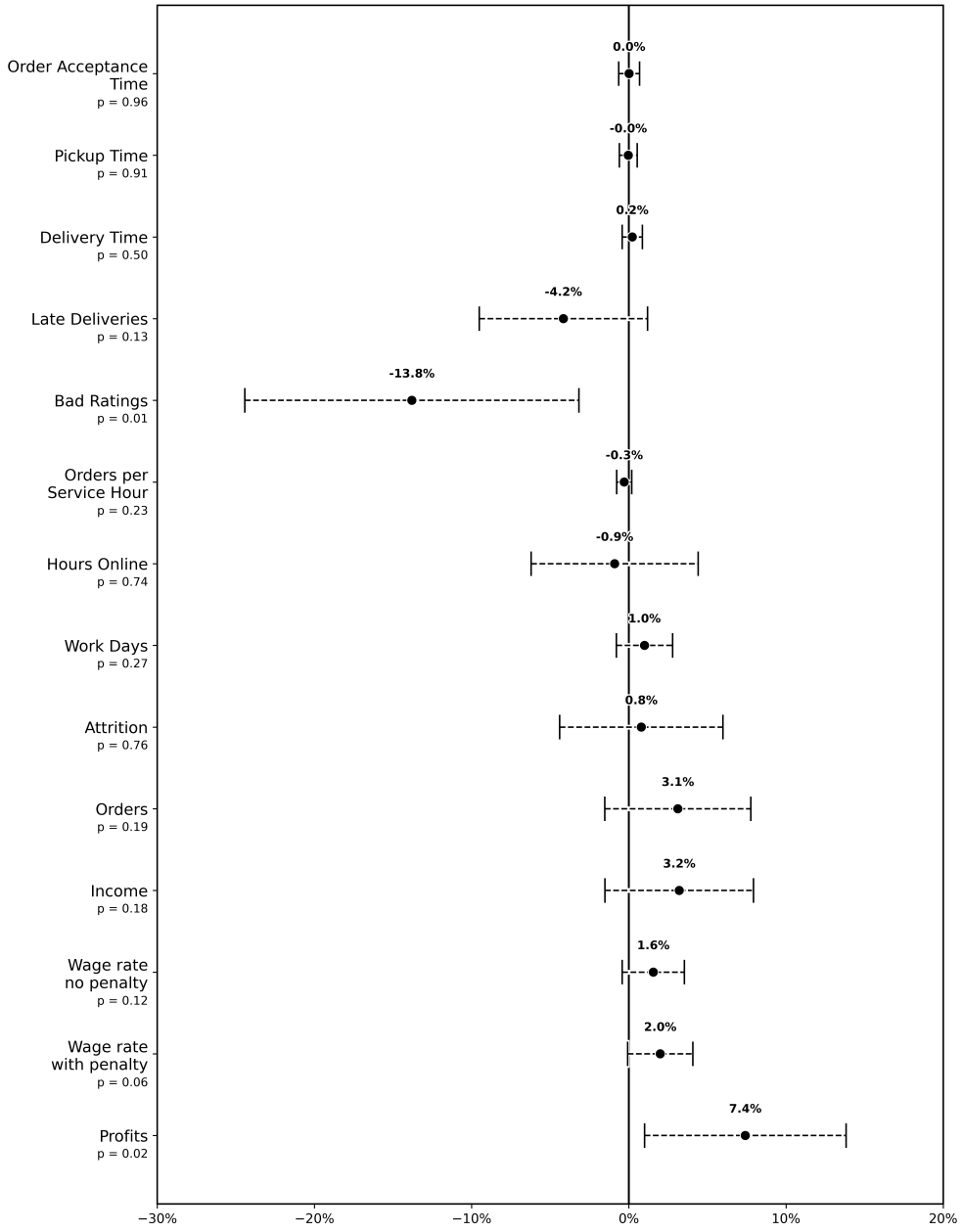
Notes: This figure presents heterogeneous treatment effects of the AI tool by pre-AI worker tenure. For each outcome, we estimate our main DiD model separately for workers within that group. The p-values shown are tests for equality of treatment effects between groups and are estimated using a standard triple-interaction model (including all three double interactions). The p-value is a test of whether $DHH_i \times Post_t \times z_i = 0$, where z_i is the heterogeneity characteristic of interest, in this case, pre-AI tenure. Pre-AI tenure is categorized as Low or High using a median split across workers, with tenure measured as of the date of AI implementation, t_0. The sample consists of workers who are present at t_0. There are observations from 482 DHH workers and 2,995 non-disabled workers.

Figure A.12: Heterogeneity in Effects of AI Tool by Pre-AI Productivity



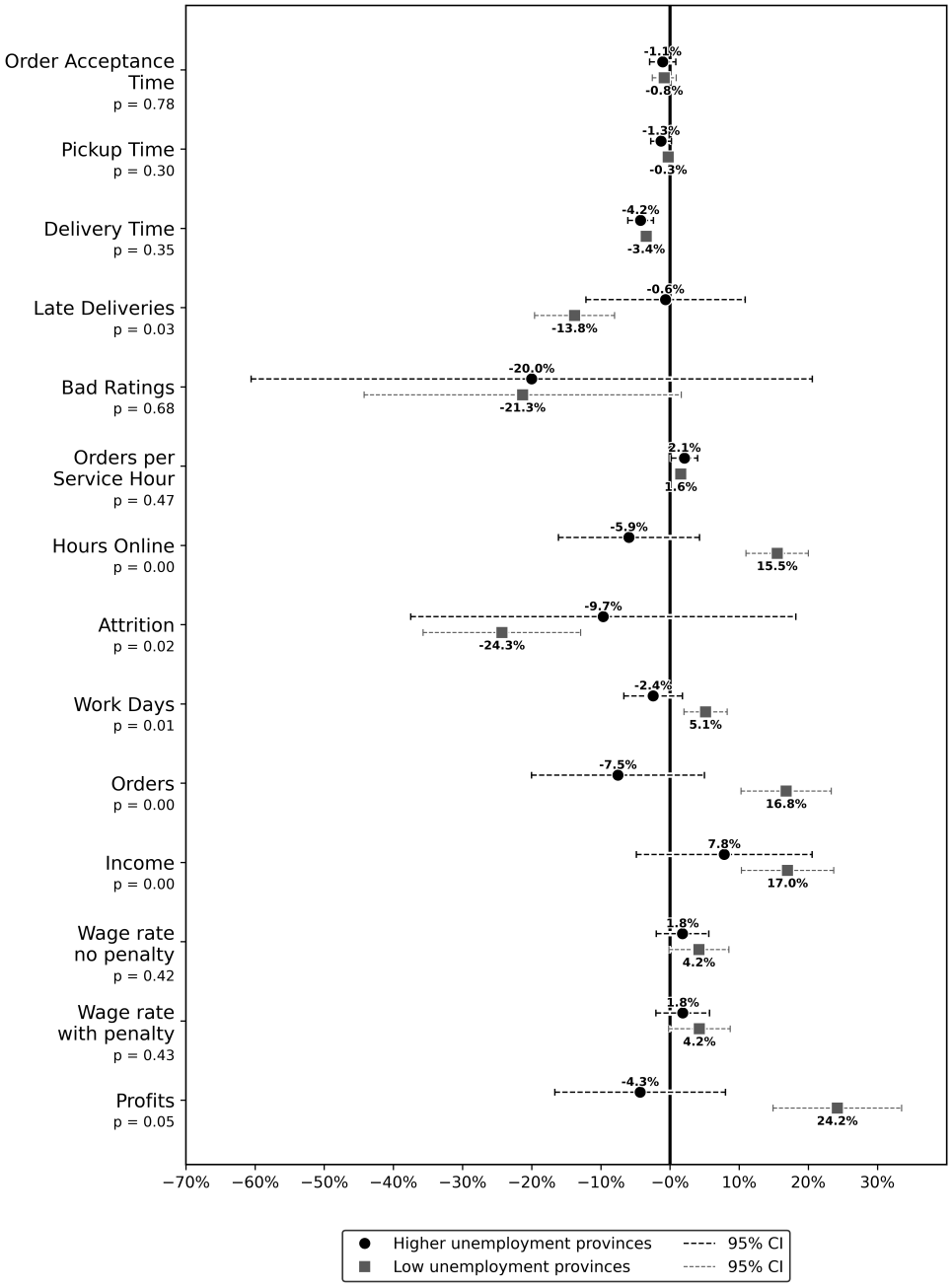
Notes: This figure is similar to Figure A.11, but splits instead of pre-AI productivity. To construct the productivity measure, we first estimate worker fixed effects from a regression of the inverse hyperbolic sine of weekly orders on worker and year-week fixed effects, using only pre-AI observations. We then regress these estimated worker fixed effects on an indicator for team (versus gig) configuration, so that the resulting residuals capture productivity variation net of systematic differences between team and gig workers. Workers are classified as high- (low-) productivity if their residual falls above (below) the sample median. The sample consists of worker-months from 1,515 DHH and 15,397 non-disabled workers.

Figure A.13: Robustness on Heterogeneity in the Effects of AI Tool by Disability Severity: Dosage Response Model



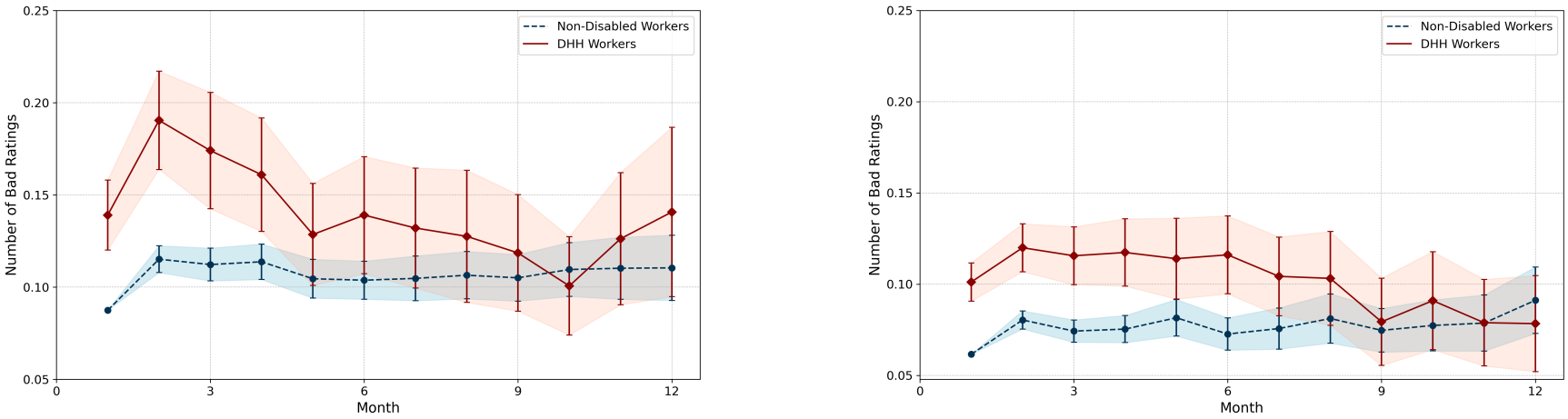
Notes: This figure is similar to Figure 9 in the main text. The difference is that instead of splitting up DHH workers by profound and non-profound hearing loss, we estimate a dosage response model. Each dot shows the coefficient on the dosage response coefficient the p-value is a test of whether it is zero. The original Level 1 (most severe, $>$ 90 dB hearing loss) to Level 4 (least severe, 41–60 dB) is transformed using Severity $= 5 -$ Level. The figure plots the coefficient on the triple interaction of $DHH_i \times Post_t \times \text{Severity}_i$. Whiskers indicate 95% confidence intervals clustered by driver. The sample consists of worker-months from 3,982 DHH and 36,719 non-disabled workers.

Figure A.14: Heterogeneity in the Effects of the AI Tool by Provincial Unemployment Level



Notes: This figure presents heterogeneous treatment effects by provincial unemployment level. It is similar to Figure A.11 in the main text, but instead splits the sample by provincial unemployment rate. Provinces are divided into high- and low-unemployment groups using the median unemployment rate as the cutoff. The sample includes 31 provincial-level administrative units ("provinces") in mainland China. Unlike for other heterogeneity analyses, standard errors are clustered by province, reflecting that we study a characteristic that varies by province. The sample consists of worker-months from 3,982 DHH and 36,719 non-disabled workers. The sample size is smaller than in our main analyses because province is missing for some workers. Information on unemployment is from the 2021 China Labour Statistical Yearbook. Lower-unemployment provinces are Qinghai, Xinjiang, Shanxi, Hunan, Jiangsu, Anhui, Guangdong, Guangxi, Zhejiang, Tibet, Shanghai, Jiangxi, Shandong, Chongqing, and Hubei. Higher-unemployment provinces are Hebei, Hainan, Beijing, Heilongjiang, Jilin, Fujian, Henan, Gansu, Shaanxi, Sichuan, Tianjin, Inner Mongolia, Yunnan, Ningxia, Liaoning, and Guizhou.

Figure A.15: Tenure-Productivity Curves in Bad Reviews, Before and After the AI Tool



(a) Pre-AI tool

(b) Post-AI tool

Notes: This figure is similar to Figure 7, but the outcome is number of bad ratings. Error bars show 95 percent confidence intervals accounting for clustering by worker. The samples in each panel are the same as in Figure 7.

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 DHH and Non-DHH Workers

	(1) Baseline	(2) Only Gig	(3) Only Team
Panel A: Efficiency			
Order Acceptance Time	0.034**	0.081***	-0.020
	(0.016)	(0.030)	(0.014)
Pickup Time	-0.006	-0.014	-0.001
	(0.009)	(0.013)	(0.013)
Delivery Time	0.045***	0.022	0.055***
	(0.011)	(0.014)	(0.015)
Late Deliveries	0.087**	0.107**	0.095*
	(0.035)	(0.048)	(0.053)
Bad Ratings	0.313***	0.310***	0.351***
	(0.058)	(0.081)	(0.083)
Orders per Service Hour	-0.019**	-0.017**	-0.015
	(0.009)	(0.008)	(0.015)
Panel B: Labor Supply			
Hours Online	0.049	0.045	0.060*
	(0.044)	(0.049)	(0.034)
Work Days	0.043***	0.024	0.051***
	(0.010)	(0.019)	(0.008)
Attrition	-0.414***	-0.284***	-0.630***
	(0.041)	(0.047)	(0.071)
Panel C: Total Output			
Orders	0.053**	0.007	0.087***
	(0.026)	(0.048)	(0.024)
Panel D: Income, Wages, and Profits			
Income	0.054**	0.007	0.089***
	(0.026)	(0.048)	(0.026)
Wage rate no penalty	-0.094***	-0.107***	0.012
	(0.020)	(0.022)	(0.020)
Wage rate with penalty	-0.102***	-0.115***	0.010
	(0.020)	(0.022)	(0.020)
Profits	0.118***	0.098	0.112***
	(0.040)	(0.073)	(0.032)

Notes: Standard errors clustered by worker in parentheses. Each coefficient here is from a separate regression. Each regression follows the procedure in Section 2.1. For all outcomes besides late deliveries, bad ratings, and attrition, we use an inverse-hyperbolic sine transform on y. For late deliveries, bad ratings, and attrition, the coefficients shown here are the OLS regression coefficients divided by the pre-AI DHH mean, so that the coefficients here can be interpreted as percentage differences. 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
Panel A: Efficiency							
Order Acceptance Time	-0.008**	-0.010*	0.003	-0.010*	-0.006	-0.001	-0.009**
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Pickup Time	-0.004	0.002	-0.006	-0.005	-0.006	0.001	-0.005
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Delivery Time	-0.035***	-0.025***	-0.040***	-0.051***	-0.035***	-0.026***	-0.037***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)
Late Deliveries	-0.112***	-0.019	-0.193***	-0.148***	-0.135***	-0.011	-0.112***
	(0.038)	(0.055)	(0.055)	(0.044)	(0.045)	(0.036)	(0.038)
Bad Ratings	-0.218***	-0.126	-0.269***	-0.252***	-0.197***	-0.086	-0.218***
	(0.068)	(0.110)	(0.088)	(0.083)	(0.063)	(0.073)	(0.068)
Orders per Service Hour	0.016***	0.009*	0.016***	0.021***	0.015***	0.009**	0.017***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Panel B: Labor Supply							
Hours Online	0.106***	0.133**	0.072***	0.136***	0.100***	0.106***	0.106***
	(0.034)	(0.056)	(0.025)	(0.046)	(0.035)	(0.034)	(0.034)
Work Days	0.035***	0.069***	0.017	0.040***	0.033***	0.051***	0.035***
	(0.012)	(0.026)	(0.011)	(0.015)	(0.012)	(0.015)	(0.012)
Attrition	-0.210***	-0.188***	-0.125	-0.193***	N.A.	-0.119**	-0.210***
	(0.052)	(0.067)	(0.089)	(0.054)	N.A.	(0.053)	(0.052)
Panel C: Total Output							
Orders	0.105***	0.151**	0.077**	0.129***	0.101***	0.128***	0.105***
	(0.031)	(0.062)	(0.032)	(0.038)	(0.031)	(0.036)	(0.031)
Panel D: Income, Wages, and Profits							
Income	0.106***	0.152**	0.078**	0.132***	0.102***	0.130***	0.106***
	(0.031)	(0.062)	(0.033)	(0.039)	(0.032)	(0.036)	(0.031)
Wage rate no penalty	0.036***	0.025	0.046***	0.033*	0.035**	0.036***	0.036***
	(0.014)	(0.022)	(0.015)	(0.019)	(0.014)	(0.014)	(0.014)
Wage rate with penalty	0.036**	0.026	0.046***	0.027	0.034**	0.036**	0.036**
	(0.014)	(0.023)	(0.016)	(0.019)	(0.015)	(0.014)	(0.014)
Profits	0.168***	0.183**	0.140***	0.244***	0.184***	0.200***	0.168***
	(0.043)	(0.092)	(0.039)	(0.051)	(0.040)	(0.058)	(0.043)
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 clustered by worker 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 (distance between restaurant and customer) included.

Table A.5: Robustness Checks for the Effect of AI Tool on Bad Ratings

	(1)	(2)	(3)	(4)
Panel A: Pre-AI Means and OLS Specifications				
	Pre-AI mean bad ratings per week, non-DHH workers	Pre-AI mean bad ratings per week, DHH workers	OLS, unnormalized	OLS, normalized relative to pre-AI DHH mean
Mean (Pre-AI) or $DHH \times Post$	0.10	0.15	-0.032^{***}	-0.218^{***}
			(0.010)	(0.068)
Panel B: Dependent Variable = Bad Ratings per Week				
	Baseline: OLS, normalized relative to pre-AI DHH mean	Poisson	OLS, normalized controlling for completed orders	
$DHH \times Post$	-0.218^{***}	-0.185^{***}	-0.239^{***}	
	(0.068)	(0.071)	(0.068)	
Panel C: Dependent Variable = Bad Ratings per 10,000 Orders				
	OLS, normalized relative to pre-AI DHH mean	OLS, order-weighted		
$DHH \times Post$	-2.350^{**}	-2.197^{***}		
	(1.024)	(0.568)		

Notes: Each coefficient is from a difference-in-differences specification using the same fixed effects as in the main text. Standard errors are reported in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A.6: Robustness Checks for the Effect of AI Tool on Late Deliveries

	(1)	(2)	(3)	(4)
Panel A: Pre-AI Means and OLS Specifications				
	Pre-AI mean late deliveries per week, non-DHH workers	Pre-AI mean late deliveries per week, DHH workers	OLS, unnormalized	OLS, normalized relative to pre-AI DHH mean
Mean (Pre-AI) or $DHH \times Post$	4.79	5.11	-0.853^{*}	-0.167^{*}
			(0.497)	(0.097)
Panel B: Dependent Variable = Late Deliveries per Week				
	OLS, normalized relative to pre-AI DHH mean	Poisson	OLS, normalized controlling for completed orders	
$DHH \times Post$	-0.167^{*}	-0.164^{*}	-0.255^{**}	
	(0.097)	(0.094)	(0.103)	
Panel C: Dependent Variable = Late Deliveries per 10,000 Orders				
	Baseline: OLS, normalized relative to pre-AI DHH mean	OLS, order-weighted normalized relative to pre-AI DHH mean		
$DHH \times Post$	-0.112^{***}	-0.089^{**}		
	(0.038)	(0.038)		

Notes: Each coefficient is from a difference-in-differences specification using the same fixed effects as in the main text. Standard errors are reported in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

B Impact of AI Tool on Good Ratings of DHH Workers

This section examines how the introduction of the AI tool affects the number of good ratings received by DHH workers. In the main text, we focus on bad ratings because those are more directly relevant to the platform and have monetary consequences for workers. By contrast, based on our interviews with the platform, good ratings do not generate direct practical benefits for workers, although they may provide psychological encouragement or a sense of appreciation.

We apply the same empirical framework used in the main text: column (1) reports the pre-AI difference specification from equation (2), and column (2) reports the overall treatment effect using the difference-in-differences specification from equation (3). Table B.1 presents the results.

Table B.1: Pre-AI Differences and Impact of AI Tool on Good Ratings of DHH Workers

	(1)	(2)
	Pre-AI Differences	DiD Results
DHH_i	0.165***	
	(0.038)	
$DHH_i \times Post_t$		0.006
		(0.051)
Observations	180,375	480,070
R^2	0.206	0.692

Notes: Column (1) reports pre-AI differences in good ratings between DHH and non-disabled workers. Column (2) reports difference-in-differences estimates of the impact of the AI tool on good ratings for DHH workers relative to non-disabled workers. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) shows that, prior to the introduction of the AI tool, DHH workers received significantly more good ratings than non-DHH workers. The estimate implies that DHH workers received 16.5% more good ratings than non-disabled workers in the pre-AI period. In comparison, our main results on bad ratings show that DHH workers also received 31.3% more bad ratings before the AI tool was introduced. Together, these findings suggest that DHH workers received more customer feedback overall, both positive and negative, than non-disabled workers prior to the intervention. Column (2) shows that the AI tool did not have a statistically significant effect on good ratings for DHH workers relative to non-DHH workers.

Taken together, these results suggest that while DHH workers were already receiving more good ratings prior to the AI tool, the tool itself did not further improve this dimension

of customer feedback. This contrasts with our findings for bad ratings in the main text, where the AI tool meaningfully reduced negative feedback. One interpretation is that the AI tool mainly operates by reducing downside communication failures, rather than shifting the upper tail of customer satisfaction.

C 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):

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

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

	(1)
$DHH_g \times Post_t$	0.275**
	(0.110)
Observations	312
R-squared	0.939

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 C.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.

D Qualitative Data

D.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 a group interview in Summer 2024 with seven DHH workers, all team workers from a station in Hangzhou. Workers were recruited via the station manager asking for volunteers to get information about difficulties faceed by DHH workers in delivering food. To facilitate communication with DHH workers, we hired a translator proficient in sign language. Workers were seated around a long table. Two coauthors took notes, along with representatives of the Platform. Workers received compensation from the Platform, around 300-400 CNY for their time.

Table D.1 summarizes the demographics of the workers interviewed. As shown in Table 1, most workers on the Platform, both DHH and non-disabled, are men. Our interview sample also consists entirely of men, but it is somewhat younger than the overall DHH worker population on the Platform. We therefore view it as broadly similar in gender composition, while noting that it skews younger in age.

Table D.1: Summary of Interviewee Demographics

Age	Hometown	Tenure on Platform	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

Below, we list the five main questions asked, as well as summarize the answers of the 7 DHH worker respondents. These are based on the notes taken by coauthors.

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 worker 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.

During the interviews, a representative of the Platform asked if workers were aware of the AI-based IOC tool and did they use it, and all workers answered Yes to both questions.

D.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 D.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. This hands-on approach enabled us to closely observe and document the workers' daily routines and the specific difficulties they encounter.

In total, we conducted fieldwork alongside three DHH workers. The first was a pilot, during which we worked for around an hour as a starting point to get to know the routines. During the pilot, we observed that upon arrival at, or near, the delivery destination, the worker usually used the AI function to place a call to the customer. Sometimes the screen displayed a notification that the customer did not answer the call; in those situations, we observed the worker repeatedly using the function "initiate intelligent outbound call" to send the message again. In cases where there were multiple orders to be delivered to one place, the worker could use the AI function to simultaneously send out intelligent outbound calls for multiple orders. Below, we document in detail the two full-day fieldwork sessions that we conducted.

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 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

Figure D.1: Immersive Fieldwork with Delivery Workers



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



(b) Author observing a DHH worker during deliveries

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 D.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 [chengzhongcun in Mandarin], but the address was inaccurate. Upon arriving, the worker used the intelligent outbound call to contact the customer, who messaged back, "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, so 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:

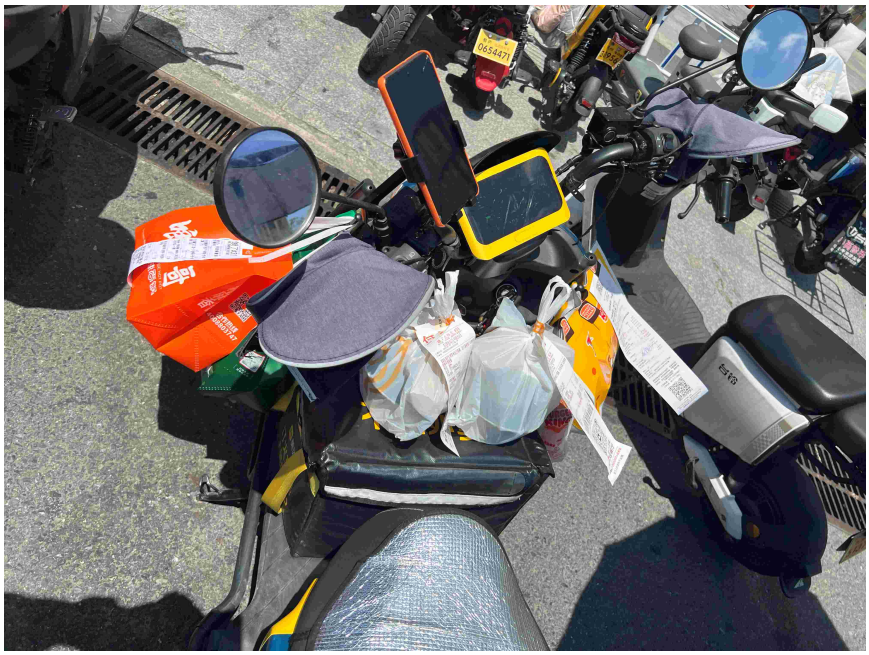
    - Workers must perform three actions during the delivery process: 'confirm arrival at the store,' 'confirm food pickup,' and 'confirm delivery.' The app only allows workers to confirm arrival at store within a certain range of the store. However, if the restaurant has not prepared the food within three 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 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.
    - 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 D.2: Essential Equipment for Outdoor Work



Notes: Figure D.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 D.3: Concurrent Orders During Peak Hours



Notes: Figure D.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.

E Further Information on Hearing Loss Levels

The following is an English translation of Section 5.3 of (GB/T 26341-2010), performed by ChatGPT on January 25, 2026:

- *5.3.2 Hearing Disability Level I. The structure and function of the auditory system are extremely severely impaired. The average hearing loss of the better ear is greater than 90 dB HL. Speech communication cannot be carried out relying on hearing. There are extremely severe limitations in activities such as understanding and communication, and extremely serious barriers exist in participation in social life.*

- *5.3.3 Hearing Disability Level II. The structure and function of the auditory system are severely impaired. The average hearing loss of the better ear is between 81 and 90 dB HL. There are severe limitations in activities such as understanding and communication, and serious barriers exist in participation in social life.*

- *5.3.4 Hearing Disability Level III. The structure and function of the auditory system are moderately to severely impaired. The average hearing loss of the better ear is between 61 and 80 dB HL. There are moderate limitations in activities such as understanding and communication, and moderate barriers exist in participation in social life.*

- *5.3.5 Hearing Disability Level IV. The structure and function of the auditory system are moderately impaired.The average hearing loss of the better ear is between 41 and 60 dB HL. There are mild limitations in activities such as understanding and communication, and mild barriers exist in participation in social life.*

Drivers with hearing loss of 40 dB or less are not issued a certificate of disability.

F Potential Nondisclosure of DHH Status

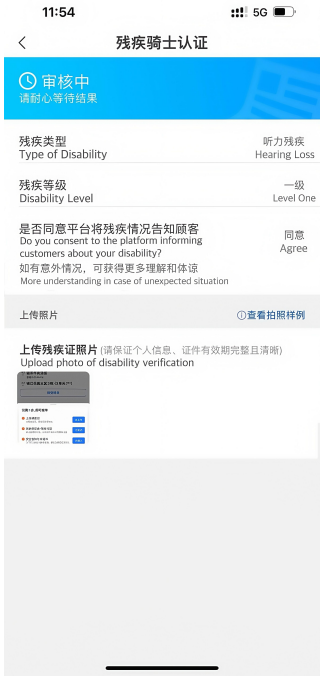
This Appendix discusses the issue of potential nondisclosure of DHH status by workers.

Why workers disclose. As noted in Section 1 of the main text, based on our qualitative work and understanding of the Platform, essentially all certified DHH workers disclose their disability to the Platform, for two main reasons.

First, it is difficult to hide, especially for team workers. In the team work model, drivers have managers, and a worker's manager would need to know if they are DHH. Drivers in both work models (gig and team) also generally receive calls from customer service or customers, and the Platform must decide whether those calls should go through.

Second, DHH workers have a strong personal incentive to disclose their disability so that the Platform can accommodate them. People with disabilities in China are accustomed to disclosing their disability to employers and the government in order to receive potential accommodations. For example, if DHH workers do not disclose their status to the Platform, customers cannot be informed of their disability and may be less understanding in unexpected situations. Figure F.1 shows the screen from the registration process where workers can disclose their DHH status.

Figure F.1: Screen During Registration Process where Workers Indicate Disability Status



Notes: This figure shows the original text, as well as text translated into English.

Comparison of severity between our data and the broader population. Table F.1 compares the distribution of hearing loss severity in our Platform sample to that among registered disability certificate holders in China. Workers with profound hearing loss (Level 1) account for 54% of DHH workers on the Platform, compared to 26% among certificate holders overall in 2012.[1]

Table F.1: Distribution of Hearing Loss Severity

Source	Level 1	Level 2	Level 3	Level 4
Registered certificate holders (2012)	26%	24%	23%	26%
Our Platform sample	54%	16%	14%	16%

Notes: This table compares disability severity prevalence between the overall Chinese population and our data, which covers new DHH workers on the Platform between 2021-2023.

In principle, an analyst might try to use quantitative tests to detect nondisclosure of DHH status on the Platform, but this is extremely difficult. One might compare the share of DHH workers in our data to population shares, overall or by severity. However, such

[1]The 2012 data are from the *China Statistical Yearbook on the Work for Persons with Disabilities*. In more recent publications, severity levels are reported for all disability types combined rather than separately for hearing loss.

comparisons also reflect selection of DHH workers onto the Platform. While the share of drivers with profound HL is higher in our sample than in the population, we believe this primarily reflects selection based on outside options. As in many countries, DHH workers face substantial discrimination and barriers to employment, particularly those with profound HL, leading many to gravitate toward gig work.

Implications of potential nondisclosure. How would any nondisclosure of DHH status affect interpretation of our results? Any nondisclosure that is fixed over time (presumably the most natural kind) would not affect the main DiD results on the impact of the AI tool, since these specifications include worker fixed effects. It would only affect the pre-AI comparisons, potentially making DHH workers appear less productive if higher-performing DHH workers were less likely to disclose their status.[2]

A different issue would arise if the AI tool itself affected disclosure of DHH status. While this is theoretically possible, none of our qualitative data or conversations with the Platform suggest that the AI tool made workers more likely to disclose their disability.[3] If workers chose to disclose their DHH status following a negative shock, this would work against the observed AI effects.

G Details on the AI Tool: Intelligent Outbound Call

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 that appears to originate from the worker's phone number.

Panel (a) of Figure G.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.

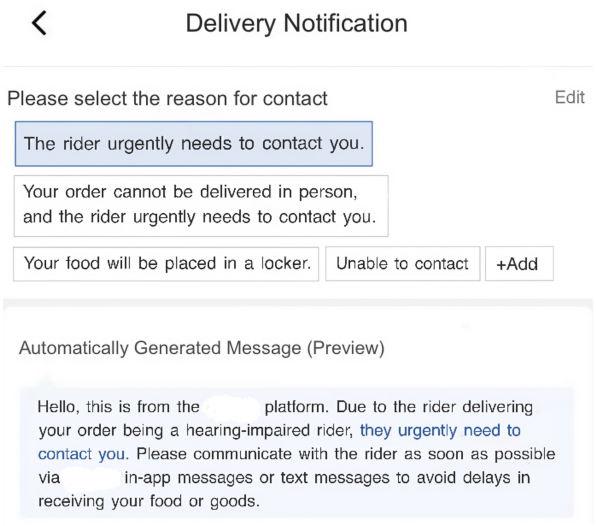
For situations beyond standard prompts, workers have the flexibility to create and store personalized messages. Panel (b) of Figure G.1 illustrates common customized messages tailored to real-world delivery challenges, such as access restrictions, security issues, or polite customer engagement.

The IOC feature significantly enhances the delivery experience in three key ways:

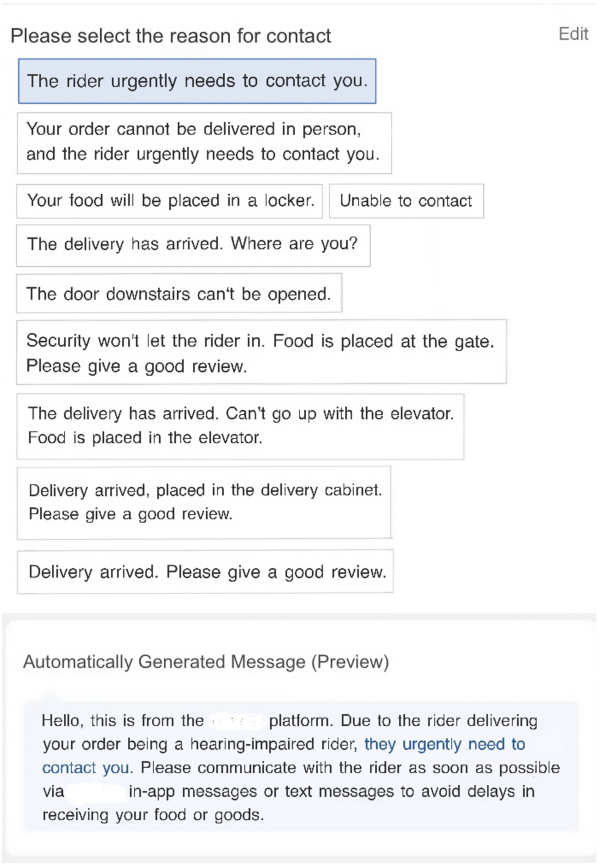
[2] That is, it would work against finding that DHH do well for the Platform in terms of output and profits.

[3] In our data, DHH status is measured as of a person's last record in the dataset. If someone disclosed their DHH status as a result of the AI tool, they would be classified as DHH for the full period of our data. Our effects are identified from within-person changes in outcomes for DHH workers.

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



(a) Default Prompts



(b) Customized Prompts

Notes: Figure G.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.

- 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.