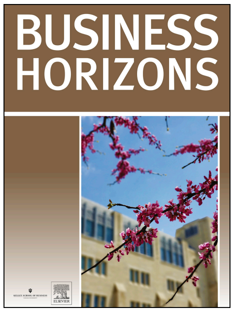
Journal Pre-proof



Fairness in crowdwork: Making the human AI supply chain more humane

Martin Gonzalez-Cabello, Auyon Siddiq, Charles J. Corbett, Catherine Hu

PII: S0007-6813(24)00126-5

DOI: https://doi.org/10.1016/j.bushor.2024.09.003

Reference: BUSHOR 2008

To appear in: *Business Horizons*

Please cite this article as: Gonzalez-Cabello M., Siddiq A., Corbett C.J. & Hu C., Fairness in crowdwork: Making the human AI supply chain more humane, *Business Horizons*, https://doi.org/10.1016/j.bushor.2024.09.003.

Fairness in crowdwork:
Making the human AI supply chain more humane

Martin Gonzalez-Cabello [a]
martin.gonzalez.cabello.phd@anderson.ucla.edu

Auyon Siddiq [a]
auyon.siddiq@anderson.ucla.edu

Charles J. Corbett [a, *]
charles.corbett@anderson.ucla.edu

Catherine Hu [b]
catherinehuihu@gmail.com

[a] Anderson School of Management, UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, USA

[b] Department of Psychology, UCLA, 1285 Psychology Building, Box 951563, Los Angeles, CA 90095-1563, USA

*Corresponding author

Fairness in crowdwork:
Making the human AI supply chain more humane

Abstract

The vast quantities of data required to build artificial intelligence (AI) technologies are often annotated and processed manually, making human labor a critical component of the AI supply chain. These workers are sourced through digital labor ("crowdwork") platforms that often are unregulated and offer low wages, raising concerns about labor standards in AI development. Using the results of a survey, this article aims to shed light on the experiences and perceptions of fair treatment among workers in the AI supply chain. The study reveals significant variability in workers' experiences, identifies potential drivers of fairness, and highlights how design choices by labor platforms can significantly affect worker welfare. Drawing on lessons from physical supply chains, the article offers practical guidance to managers on how to enhance worker welfare within the AI supply chain and ensure that AI technologies are responsibly sourced.

KEYWORDS:

1. The human supply chain behind AI and crowdwork

In the supply chains for garments, electronics, and other physical goods, the process is often broken down into increasingly small and standardized tasks. This makes them easier to outsource and offshore to geographically distant locales and through one or more layers of intermediaries. Downstream firms end up with limited visibility into and diminished control over upstream labor practices. At the same time, such work attracts individuals with few other employment opportunities. Together, this creates conditions ripe for mistreatment of workers, which is well-studied in the management literature (eg., Aydin et al., 2014).

Supply chains for digital goods display similar trends. Information products also involve chains (Karmarkar et al., 2007). As communication technology improved, software development was outsourced to India; small repetitive tasks, such as insurance paperwork and airline ticket processing were initially physically shipped and later digitized and outsourced to locations with lower labor costs such as the Caribbean and the Philippines (Metters et al., 2008). More recently, spurred by the growth in artificial intelligence (AI), "crowdwork" platforms have emerged to facilitate outsourcing of digital tasks – including data annotation, audio transcriptions, and image tagging – to workers anywhere in the world (Berg et al. 2018). As with physical supply chains, the downstream firms that rely on the data processed through crowdwork have minimal visibility into upstream labor standards, and these platforms also often attract individuals with limited opportunities, again creating conditions conducive to abusive practices. Emerging evidence indicates that crowdworkers often experience poor working conditions, such as low pay, unfair rejection of their work, and limited ability to take breaks. Gray and Suri (2019) refer to the manual labor behind AI as "ghost work" owing to its low visibility to end users and the weak protections afforded to crowdworkers. Figure 1 illustrates that the digital supply chains behind AI possess several characteristics similar to the complex global supply chains for physical goods, and hence suffer from some of the same challenges.

Despite these parallels, there are important differences. First, manufacturing work is more likely to be an individual's primary source of income, as it is more difficult to do part-time; crowdwork is fully digital, often done remotely, and compensated per-task, all of which makes it easier to work part-time. Second, it is harder for full-time manufacturing workers to frequently switch jobs or work for multiple employers, whereas crowdwork allows individuals to be active on multiple platforms simultaneously. Third, firms with poor working conditions may prevent workers from organizing or forming unions; platforms cannot entirely prevent crowdworkers from finding and communicating with each other.

[Insert Figure 1 About Here]

For physical supply chains, issues related to labor conditions are by now relatively well understood, although far from resolved. Downstream firms are sometimes held accountable by consumers and regulators for labor conditions upstream; they rely on audits, certification schemes, enhanced transparency, longer-term contracts with vetted suppliers, and other mechanisms to satisfy those demands. Digital supply chains are less well understood, and their working conditions far less regulated. However, their digital nature provides researchers a unique opportunity to communicate directly with crowdworkers, to understand the challenges they face and identify opportunities to improve their well-being.

All this raises several questions about whether crowdworkers are treated fairly and how that varies across workers and platforms. Specifically, we ask:

1. How are crowdworkers' perceptions of fair treatment by platforms shaped by their work experiences?

2. How do crowdworkers' experiences and perceptions of fair treatment vary between full-time and part-time crowdworkers?

3. How do crowdworkers' experiences and perceptions of fair treatment vary across platforms?

Based on a survey of crowdworkers we find that several factors are associated with fair treatment, including finding work engaging and not spending too much time looking for work. Full-time crowdworkers report worse experiences than part-timers, but believe they are treated more fairly, consistent with having fewer alternative opportunities and being more vulnerable to exploitation. Finally, some platforms are seen as much more fair than others. Altogether, the AI supply chain should explore strategies adopted in physical supply chains to ensure ethical working conditions, while also leveraging the ways in which digital supply chains differ.

We focus on fairness because of its relevance and significance in the context of crowdwork (Fieseler et al., 2019) and the importance of fair labor standards more broadly. The crowdwork industry is rapidly expanding: One estimate puts the total number at 9 million (Berg, 2015), while another has 4.8 million crowdworkers in seven countries in Africa alone (Anwar & Graham, 2020). A major driver is the enormous appetite of AI systems for data (whether images, speech, text, or other formats) that needs to be curated and labeled (Vaughan, 2018).

Not all AI relies on crowdwork, and not all crowdwork feeds into AI, but the two are inextricably linked. As firms embed AI in their services, understanding whether the data that powers their AI is sourced ethically becomes important. An executive at a major AI-focused crowdwork platform confirmed that many downstream firms care about fairness. Friedland and Balkin (2023) point out that crowdwork platforms are vulnerable to reputational damage if customers perceive them as behaving irresponsibly towards workers. To that end, our first question is about factors that contribute to perceptions of fair treatment on crowdwork platforms.

In other parts of the gig economy, substantial differences exist between those who depend on it as their main income and those who use it as a supplemental source. Schor et al. (2020) find that supplemental earners are more satisfied and get higher earnings overall. The on-demand and ad-hoc nature of gig work provides desirable flexibility for supplemental earners, but worsens precarity for those further down the socioeconomic ladder. The uncertainty in wages and availability of work can exacerbate inequities, prompting our second question, about differences between full-timers' and part-timers' perceptions of how fairly they are treated.

Lastly, several studies document that workers on different platforms have divergent experiences (Berg et al., 2018; Rani & Furrer, 2021), possibly due to platform design and operation.

Therefore, the third question investigates whether crowdworkers perceive some platforms as more fair than others. To supplement our survey, we had several conversations with a manager at the platform that was perceived as most fair, to understand what it did differently.

By shedding light on the experiences of workers and drawing from lessons learned from physical supply chains, our research identifies opportunities for improving working conditions at a vital stage in the AI supply chain. We conclude with recommendations for how managers can improve labor standards in digital supply chains more broadly.

2. Background: Crowdwork and AI

2.1. AI's reliance on crowdwork labor

During the development of AI algorithms, large quantities of data must be organized, annotated, and processed, often by human workers. To endow a self-driving car with human-like vision, the AI must be fed millions of images, each labeled with its contents (e.g., pedestrians, stop signs, cyclists). Labeled images can help AI detect potential cancer from diagnostic scans, and voice-activated AI assistants such as Amazon's Alexa can recognize human speech thanks to troves of human-transcribed audio clips. These vast quantities of labeled training data assembled through brute force by human workers provides the "ground-truth" from which AI algorithms learn. Even after deployment, human oversight is needed; a social media feed curated by AI may require continuous moderation by humans to filter out harmful content.

Crowdwork is a way to leverage human intelligence to perform tasks that cannot be accurately completed through automation alone. Kietzmann (2017, p.152) defines crowdsourcing as "The use of IT to outsource any organizational function to a strategically defined population of human and non-human actors in the form of an open call." Prpić et al. (2015) distinguish several types; we focus on what they refer to ask micro-task crowdsourcing. Amazon launched Mechanical Turk (MTurk) in 2005 as a way to complete small jobs that are difficult for computers, but simple for humans. Employers post tasks, set payments, and request a specific number of workers. Workers select which tasks to complete. MTurk paved the way for numerous other platforms. Some (e.g., Toloka, Appen, Prolific) function as open marketplaces similar to MTurk, and generate revenue through service fees, while others (e.g., Scale AI) are more proactive intermediaries, charging clients per task and recruiting a distributed workforce directly. Schmidt and Jettinghoff (2016) offer an early overview of such platforms with recommendations for how firms can best engage with them.

2.2. Challenges in crowdwork

The well-being of crowdworkers has recently drawn significant attention in the academic literature and the popular press. Legal scholars argue that crowdwork platforms operate in a gray area outside the protections of employment law (Felstiner, 2011). The median wage of crowdworkers is estimated to be \$2 to \$4 per hour (Berg et al., 2018; Horton & Chilton, 2010), and many rely on it as their main source of income. Workers are typically paid per-task, so time spent searching for tasks lowers their effective hourly wage. In a study on MTurk the average requester pays over \$11/hour, but lower-paying requesters post much more work; once unpaid work is accounted for—including time spent searching for tasks or work that is eventually rejected without pay—the effective median hourly wage is closer to $2 per hour (Hara et al., 2018). Our survey respondents complained about low wages: “I don’t like how little work there is on Amazon Mechanical Turk these days. Too many people joined in, and now we’re all just scrapping for pennies”, and “I don’t like days when MTurk is slow and thousands of people are trying to catch a few good paying jobs”.

While this modest income may be relatively attractive in developing countries, concerns persist about the precarity and fair treatment of crowdworkers (Anwar & Graham, 2020), mirroring the broader gig economy (Schor et al., 2020). A typical day for a data-labeler consists of clicking through thousands of images, and content moderators may be subject to images of violence, pornography, and other graphic content, with little consideration of their psychological well-being (Gray & Suri, 2019; Roberts, 2019). Recently, it came to light that OpenAI, the creator of the chatbot ChatGPT, relied on Kenyan workers paid $2 per hour to label thousands of text passages, sometimes containing hate speech or descriptions of violence, to teach the AI to avoid reproducing toxic language (Time, 2023). Friedland and Balkin (2023) point out the reputational risks for crowdwork platforms when such scandals erupt.

Some of the challenges with crowdwork are functions of platform design: difficulty in communicating with clients, a lack of support in resolving disputes, and uncertainty in payments. On MTurk, workers are paid at the discretion of the requester, which leaves them vulnerable to exploitation and gives them little recourse (McInnis et al., 2016). Shevchuk and Shebkov (2018) report that 73% of workers on a Russian platform experienced “agreement violation” by the client. One of our respondents wrote “Mechanical Turk is very underpaid; many requestors promise bonuses but don’t pay them. Many requesters engage in wage theft in the form of unfair rejection of careful work.” Crowdwork platforms often use algorithms to coordinate work and quantify work performance, which impacts perceived conditions and meaningfulness of work (van Zoonen et al., 2023, 2024).

The quality of data obtained from crowdwork is also a concern (Wang et al., 2017; Paolacci et al., 2010; Peer et al., 2022). For physical supply chains more socially irresponsible practices lead to lower product quality (Distelhorst & McGahan, 2022). Skilled workers with other options may stop engaging in crowdwork because wages are low—leading to lower data quality—and clients pay low wages in anticipation of low quality. Since accurate training data is vital for building AI, the well-being of crowdworkers may have implications for its efficacy. This “race to the bottom” might be exacerbated by AI itself, as tools such as ChatGPT start to outperform human crowdworkers at a fraction of the cost (Wright & Schultz, 2018; Gilardi et al., 2023). Despite such developments, it is unlikely that human crowdworkers will ever become unnecessary: Shumailov et al. (2023) point to the risk of model collapse if AI models start relying too much on data generated by AI models, and Castro et al. (2023) point to the societal pitfalls that can occur, suggesting that improved human-AI interactions are one solution.

2.3. Prior survey work on crowdworkers’ well-being

Several studies conduct surveys or interviews with crowdworkers, some across multiple platforms (Berg et al., 2018; Rani & Furrer, 2021; Churchill & Craig, 2019; Lehdonvirta, 2018). None has explored how the same workers’ experiences vary across platforms, which is crucial for understanding the role of platform design in shaping worker welfare. A 2017 study by the

International Labor Organization (ILO) surveyed 3,500 crowdworkers in 75 countries working on five different platforms (Berg et al, 2018). A significant minority depended on crowdwork for income, but most lacked protections such as health insurance. A qualitative survey found that a platform's features can shape perceptions of procedural fairness (Fieseler et al., 2019). Our study uncovers potential determinants of fairness, how experiences and perceptions of fairness vary between those who depend on crowdwork for their primary income and those who don't, and how they vary with platform design.

A unique aspect of our survey is comparing workers' experiences across different platforms. In physical supply chains, surveying workers in an upstream factory is challenging even if its location is known to managers downstream. It is near-impossible to compare two factories based on the experience of a common set of workers who have worked at both. In contrast, crowdworkers can to some extent be accessed through the very platforms they work on, enabling us to identify workers with experience on multiple platforms. We find substantial differences between full-time and part-time workers, and between platforms. We conclude by revisiting initiatives in physical supply chains to ensure more ethical working conditions, and speculate on how they may be adopted or adapted for the AI supply chain.

3. What the crowdworkers told us

3.1. How we conducted our survey

We conducted our survey during Spring of 2022 using Qualtrics and administered through Amazon's Mechanical Turk (MTurk) using the CloudResearch service (Litman et al., 2017), with assistance from [blinded for peer review]. MTurk is by far the largest platform, and hence also most likely to yield enough respondents who are active on other platforms. We did not aim for our sample to be representative of all crowdworkers across all platforms, but rather to include crowdworkers who are active on multiple platforms. We took care to account for non-US working hours and to ensure fair payment. Our final sample included 934 complete responses. (More details are available from the authors.)

Our respondents ranged in age from 18 to 69 years, with a mean of 35. 47% were only active on MTurk, 26% on one additional platform, and 27% on two or more. The majority (55%) had a bachelor's degree, 65% identified as male, and respondents were roughly evenly distributed among the US, India, and the rest of the world. For 39% crowdwork was their primary source of income, slightly higher than the 32% in Berg et al. (2018).

We used six subjective measures of workers' experiences to understand the typical process they go through to find, complete, and be compensated for their work:

1. The amount of time workers spent looking for tasks to complete;

2. The proportion of tasks workers regretted accepting;

3. How frequently workers felt obligated to accept tasks;

4. How often workers had the flexibility to take breaks during or between tasks;

5. How engaging workers found the tasks to be;

6. How fairly the workers felt they were treated by the platform overall.

The survey was certified exempt by [institution blinded for peer review] IRB (IRB#22-000509). We initially focus on MTurk, and later contrast that to other platforms.

3.2. Fair play or all is fair?

While many crowdworkers considered MTurk as "somewhat fair" or "very fair", 14% found it "somewhat unfair" or "very unfair", with 6% in the latter category. Although this is a minority, 6% of the 9 million estimated crowdworkers would be half a million individuals claiming "very unfair" treatment. Moreover, some may consider a platform reasonably fair by comparison if their alternative employment options are worse, so the number actually being treated unfairly may be higher still.

We use risk ratios to quantify the extent to which each experiential metric is associated with perceptions of fairness. For instance, among users who spend over 30% of their time searching for work, 20% perceive the platform as "somewhat unfair" or "very unfair"; in contrast, among those who spend less than 30% of their time searching for work, only 11% view the platform as unfair. The risk ratio of 20/11 = 1.9 signifies that users dedicating more than 30% of their time searching for tasks are 1.9 times more likely to view the platform as unfair. One worker illustrates this: "I only work on MTurk. The thing I like [is] I can work according to my convenient timings. The thing I don't like is sometimes I have to wait hours for work".

Table 1 shows the risk ratios. Most strikingly, workers who find tasks "somewhat" or "very boring" are six times as likely to perceive the platform as unfair relative to those who find them at least "somewhat engaging". Though causality is ambiguous, this points to a strong association between crowdworkers feeling they are treated fairly and the extent to which they can ascribe meaning to their work. Workers who complete interesting, engaging tasks may feel more valued by the platform, whereas those who perform repetitive tasks they find boring or meaningless may view the platform as treating them unfairly. Workers who regret accepting tasks, spend significant time looking for tasks, feel obligated to accept tasks, and are unable to take frequent breaks are also all much more likely to view the platform as unfair; except for the last metric, all differences are statistically significant ($p < 0.05$). For our first research question, whether crowdworkers believe a platform treats them fairly is strongly correlated with several other aspects of their experience, not just direct pay rates.

[Insert Table 1 about here]

3.3. Full-timers experience the same work differently than part-timers

We will refer to the 39% of workers who reported crowdwork as their primary source of income as "full-timers." To understand how crowdwork differs between full-timers and part-timers, we examined the six experiential metrics separately for both, summarized in Figure 2. Full-timers appear to have a worse experience than part-timers. For example, 51% of full-timers spend more than 30% of their time looking for tasks, compared to only 36% of part-timers. Full-timers need to continuously search for tasks to secure their income, especially during periods when tasks are scarce. Lehdonvirta (2018) quotes a full-time worker: “I’m sleeping in front of my computer, and to be honest with you, I am having my meals in front of my PC just to take advantage of the tasks that are available…No permanent time of sleeping, because I need to check if there [are tasks] available. …I am online for almost 20 hours a day.” Full-timers also more often regret accepting tasks than part-timers, feel more obligated to accept tasks, and report having less flexibility to take breaks. Surprisingly, despite faring worse on each metric, full-time crowdworkers view the platform as more fair (and tasks as more engaging) than their part-time counterparts: 46% responded feeling treated “very fairly”, versus only 32% of part-timers.

This addresses our second research question: full-time crowdworkers experience worse conditions than part-timers, but view the platform as more fair. This may be due to alternative employment opportunities being worse or non-existent. Our finding is consistent with the concern that workers who are reliant on crowdwork might tolerate even worse working conditions and be prone to exploitation (Berg et al., 2018).

[Insert Figure 2 About Here]

3.4. Not all platforms are created equal

One unique aspect of our research is that we compare the same workers’ experience across different platforms (by asking respondents which other platforms they worked on in addition to MTurk). Workers themselves are keenly aware that platform design matters, with one respondent writing, “MTurk was not designed for the things it is used for, and is poorly maintained, at best, so it generally is a pain to deal with.” Although respondents reported working on over 20 different platforms, only three platforms had more than 100 respondents each. In what follows, we highlight one of these platforms — Prolific, with 165 respondents (18%) — because it stood out as being perceived as significantly more fair than the others.

We asked the same six questions for each platform respondents worked on. Figure 3 shows that for the 165 who work on MTurk and Prolific, the latter provides a better experience on all six metrics. The main complaint about Prolific was that there was not enough work: crowdworkers would do more on Prolific if they could. We also compared the MTurk workers who do not work on Prolific to those who do. Of the non-Prolific workers (i.e., all 934 MTurk respondents excluding the 165 who also work on Prolific), 41% perceive MTurk as very fair, against only 21% of those who also work on Prolific. Exposure to fairer platforms may lead workers to become more acutely aware of unfair treatment. This aligns with the difference between full-timers and part-timers: perceptions of fairness depend on the available alternatives.

An interview with a Prolific manager revealed several factors that help it stand out. Prolific enforces a minimum hourly wage of $8.00 for its workers (Prolific, n.d.), significantly higher than the effective wage on other platforms, most of which do not enforce minimum pay rates. Several aspects of Prolific’s design may explain the difference in the time spent looking for work. First, the platform tries to maintain an appropriate ratio of workers to tasks, thereby ensuring sufficient work and preventing excessive competition among its workers. Second, Prolific’s upfront screening process ensures workers are qualified to complete the tasks they accept, in contrast to MTurk, where workers can be screened out as part of the task itself. Third, Prolific uses a rate-limiting algorithm that prioritizes workers who have completed fewer tasks, reducing competition for available tasks and streamlining the task assignment process. A respondent confirmed the impact of these features: "I like that it enforces a minimum wage. I also like that it distributes work." Lastly, Prolific users can enable notifications for newly posted tasks. All this contributes to crowdworkers spending less time looking for work.

[Insert Figure 3 About Here]

Workers also regret accepting tasks less frequently on Prolific than MTurk. Prolific enforces strict policies for what qualifies as a valid reason for clients to reject a completed task and refuse payment to workers. On Prolific clients must provide the platform with strong evidence of non-compliance by workers to refuse payment, in contrast to MTurk where workers are paid solely at the discretion of the client. Even if a rejection is permitted on Prolific, the client must send a rejection message to the worker. In extreme cases, Prolific will mediate disputes, and if necessary, directly compensate workers. Respondents said "Prolific stands up for task workers" and "If the requester is not paying a fair wage, you can report them and the wage is adjusted. There isn't as much work but you are respected there." Contrast this to what respondents said about MTurk: "The whole rejection mechanic is twisted against workers. A single unfair rejection can have serious impact on rating and there is no real appeal process" and "I generally think [MTurk is] an awful place for workers, without any kind of safety measures."

Among workers who engage with both platforms, 46% report "always" being able to take breaks on Prolific, compared to 35% on MTurk. Prolific workers can reserve a spot in a task for a short period of time (e.g., 10 minutes), which provides some flexibility. The difference between Prolific and MTurk is relatively small for this metric: even a platform that emphasizes the well-being of crowdworkers faces challenges in fully achieving that goal.

Responding to our third research question, the design of crowdwork platforms has a significant impact on workers' experience. Current challenges in crowdwork—including low effective wage, anxiety about finding sufficient work, unfair rejection of work, and the inability to take sufficient breaks—need not be the norm as the industry continues to expand. In addition to ethical pay, design features that limit excessive competition among workers can leave them better off than a free-for-all system that breeds anxiety about finding sufficient work and pay.

4. Learning from physical supply chains

Physical supply chains show that workers who perform repetitive tasks for low wages unseen by downstream consumers are often subjected to substandard conditions. The apparel industry, for example, has a long history of poor working conditions resulting in highly publicized scandals. While crowdworkers face fewer physical risks compared to workers on a factory floor, they also suffer low wages, job insecurity, and the stress of an unregulated workplace. Crowdworkers who are regularly exposed to graphic or sensitive content may face poorly understood psychological harms (Roberts, 2019). It is worth exploring avenues for improving labor conditions in the AI supply chain, by drawing on established practices from physical supply chains, and by leveraging ways in which they are different.

4.1. Engaging the entire AI supply chain

The fair and ethical treatment of crowdworkers requires engagement from stakeholders throughout the AI supply chain. Crowdwork platforms have a responsibility to improve working conditions, and our findings suggest that they have the power to do so through judicious design of the platform. Firms throughout the AI supply chain have a responsibility to ensure that the labeling and management of data are handled by crowdwork platforms that create reasonable working conditions. The pressure to uphold ethical labor standards can come from multiple sources, including investors, regulators, end consumers of AI products, and even employees of companies developing AI technologies. Nike was famously criticized in the 1990s for relying on sweatshop labor, and initially defended itself by stating that it did not own or operate upstream factories ("we don't make shoes"). This response drew significant public backlash, ultimately prompting Nike to adopt and advocate for more responsible practices in its supply chain (Doorey, 2011), including audits of its factories and improving labor relations after implementing lean initiatives (Distelhorst et al., 2017). This is the same kind of reputational risk that Friedland and Balkin (2023) warn platforms about. Similarly to Nike's use of lean initiatives, firms using crowdwork should adopt a similar mindset when they find themselves rejecting too many submissions: they should engage with the crowdworkers to identify and eliminate the root causes of the quality problems, rather than simply rejecting the work and moving on.

Companies developing or adopting AI technology should also aim to build long-term relationships with data-labeling firms that rely on crowdwork. In global supply chains, excessive price pressure can lead to unauthorized subcontracting, leading to violations of labor standards (Caro et al., 2021). Subcontracting occurs in crowdwork too: Graham et al. (2017) report the case of a worker (A) who bid $15 to do a search engine optimization job that had been posted with a suggested price of $50; a different crowdworker (B) got the job with a bid of $23, and then offered it to the initial worker (A) for $3.50. To protect the interests of crowdworkers, managers could limit themselves to a small set of long-term vendors and refrain from constantly shopping around for the lowest cost or fastest turnaround. Firms posting work should give crowdworkers enough time to complete the task after accepting the job, so that workers can take breaks and generally organize their lives responsibly rather than be rushed to complete the task; most AI-related crowdwork is not so time-sensitive that a few hours or days will make a difference to the downstream firm, but that extra flexibility can be invaluable for crowdworkers.

Supply chain transparency is also crucial for ensuring fair treatment of workers. Transparency enables customer firms to advocate for improved standards, and increases the reputational risk for vendors who deviate from widely adopted norms. Currently the AI supply chain remains opaque. Some data-labeling firms operate under a different identity when engaging with end customers versus crowdworkers; the data-labeling giant Scale AI uses a subsidiary named Remotasks when recruiting crowdworkers. One form of transparency would be to enable workers to publicly post reviews of platforms and requesters. Merlo et al. (2018) argue that firms can benefit from being transparent about their performance, through their own website and/or through neutral third-party websites, even if some of the reviews are negative. Further, where possible, managers interested in evaluating working conditions should rely on objective measures of welfare (e.g., the ability to take breaks or the time spent looking for work), since self-reported measures of fairness can be biased by the presence or absence of outside options.

Such performance transparency about platforms and requesters helps to educate crowdworkers about their options. In our survey, workers who had been exposed to Prolific were far more critical of MTurk than workers who had not. Existence of worker forums provides some such transparency, whether firms like it or not. Leading firms can leverage this to generate discontent among workers about conditions at lagging firms.

Audits play a vital role in global supply chains, and could similarly be instrumental in crowdwork. Downstream companies could encourage data-labeling firms to comprehensively document their operations and disclose where and how their data-labeling work is conducted. Audits could include verifying how much time workers spend online, including looking for work, that they can take breaks, and that they do not have work unfairly rejected. Although auditing is challenging in all contexts, the digital nature of AI supply chains may lend itself to technological approaches to auditing that would not be viable in physical supply chains. Online platforms such as Elevate's Laborlink and the now-defunct LaborVoices (Castka et al., 2020; McGrath et al., 2021) might be particularly effective for crowdworkers who are, by definition, online. Workers could submit anonymous reports if they regret accepting work, experience unfair rejection, or are subject to other injustices. Given the vulnerability of gig workers (which includes crowdworkers), Friedland and Balkin (2023) highlight the importance of ensuring they have access to distributive and procedural justice; they suggest that platforms can leverage their customers' moral self-awareness, which involves several levels of self-reflection, to encourage them to treat the workers more fairly. The development of AI certification programs, similar to those for fair trade or ethical sourcing, could also help provide a guarantee that an AI system was developed responsibly. The Fairwork Foundation, through its five principles for evaluating online labor platforms, is one such approach (Graham et al., 2020).

One of these principles is fair representation. In physical supply chains this is difficult to enforce if factory managers do not want workers to organize, but it is much harder to prevent crowdworkers from finding and communicating with each other, as forums such as Turkopticon and MTurk Forum illustrate. Firms should encourage crowdworkers to be organized and should use these platforms to learn more about workers' concerns and about how to be better employers. Schmidt and Jettinghoff (2016) give the example of a firm that has particularly high requirements for data quality: such a firm would benefit from using these worker groups to explain their needs and ensure that their work gets done well, rather than just reject unusually many submissions and risk being blacklisted by crowdworkers as a result.

4.2. Improving crowdwork design and technology

In addition to these supply chain measures, crowdwork platforms can improve labor conditions directly through careful design of tasks. In physical assembly lines, it is difficult to completely replace humans with robots; instead firms design ways for automation to make work safer for humans, for instance by moving parts and tools to the right height for the human assembly worker to comfortably use. Firms that request data-labeling should explore ways to make the work more meaningful, less repetitive, and less stressful. In our survey, more engaging work was strongly associated with greater perception of fairness. Managers can group similar images together to help crowdworkers complete tasks faster, or deliberately provide more variety to make the work more engaging. Some are exploring gamification (Morschheuser & Hamari, 2019). Providing clearer task descriptions and instructions can help reduce the time spent searching for work. At a minimum, they should minimize technical glitches in the tasks, which are not uncommon and waste crowdworkers' time.

Additionally, managers could consider ways to limit harms to crowdworkers from graphic or hateful content, such as pre-filtering images to allow workers to make informed judgments about content without being fully exposed to it. When automation competes with human labor in physical supply chains, already-low wages get depressed even further, and AI that is competent at data-labeling could have a similar effect on crowdwork wages. Conversely, deploying AI alongside human workers might lessen the burden currently carried by crowdworkers, especially if used for the most mundane, repetitive, or harmful tasks.

Improving the welfare of crowdworkers may also improve the technology itself, which would benefit adopters of AI. Evidence from physical supply chains suggests that improved working conditions can enhance product quality. Lean manufacturing has been shown to improve both labor standards and worker productivity (Distelhorst et al., 2017). Prolific appears to produce superior data quality (Peer et al., 2022) while treating its workers more fairly. Given that AI relies critically on the accuracy of training data, investing in the well-being of crowdworkers could yield better AI technology in the long run.

4.3. Expanding the scope of ethical AI

The definition of "ethical AI" should be broadened to explicitly account for the fair treatment of crowdworkers. To date, ethical frameworks for AI focus on important societal impacts of AI deployment, including user privacy, discrimination, and bias in AI-driven predictions and decisions; most are silent on the role of crowdworkers. Explicitly acknowledging fair treatment of crowdworkers as a key component of ethical AI will raise the visibility of human workers in the AI supply chain. A related concern is the use of work by humans who do not know that their daily activities are used for training AI systems (Morreale et al., 2023).

The regulatory landscape for global supply chains has evolved considerably, and the AI supply chain may follow a similar trajectory. California's 2012 Transparency in Supply Chains Act requires companies to disclose their efforts to uphold labor standards in their supply chain. Firms in the AI supply chain might face similar scrutiny. The ongoing debate around the classification and treatment of gig workers (e.g., AB5 and Prop 22 in California) may have implications for crowdwork. Drivers on rideshare platforms do not have as much flexibility as might appear, due to the algorithmic control exerted by the platforms (Dubal, 2023; Rosenblat & Stark, 2016). Similarly, crowdwork may offer less flexibility and autonomy than is often thought, especially for those who depend on it. In light of the vulnerability of gig workers, platforms' reputations are also vulnerable (Friedland & Balkin, 2023).

5. Conclusion

Rapid advances in digital technologies, including AI, have created enormous demand for manual data-labeling and curation – tasks carried out by workers that are mostly invisible to end users. This article sheds light on the association between their experiences on crowdwork platforms, their perceptions of fair treatment, and the role of platform design. Drawing on parallels with physical supply chains, we suggest several avenues for improving labor conditions within digital supply chains. It is clear that these issues also have implications for data quality, cost, sourcing strategy, reputation, and more. Ethics in the AI supply chain cannot be treated as a standalone issue, but require an integrated perspective encompassing all these other factors, just as is the case for environmental and social sustainability in physical supply chains.

References

Anwar, M. A., & Graham, M. (2020). Digital labour at economic margins: African workers and the global information economy. *Review of African Political Economy, 47*(163), 95-105.

Aydin, G., Cattani, K., & Druehl, C. (2014). Global supply chain management. *Business Horizons*, *57*(4), 453-457.

Berg, J. (2015). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comparative Labor Law & Policy Journal, 37*, 543-576.

Berg, J., Furrer, M., Harmon, E., Rani, U., & Silberman, M. S. (2018). *Digital labour platforms and the future of work: Towards decent work in the online world*. International Labour Office.

Caro, F., Lane, L., & Sáez de Tejada Cuenca, A. (2021). Can brands claim ignorance? Unauthorized subcontracting in apparel supply chains. *Management Science, 67*(4), 2010-2028.

Castka, P., Searcy, C., & Mohr, J. (2020). Technology-enhanced auditing: Improving veracity and timeliness in social and environmental audits of supply chains. *Journal of Cleaner Production, 258*, 120773.

Castro, F., Gao, J., & Martin, S. (2023). Human-AI Interactions and Societal Pitfalls. arXiv preprint arXiv:2309.10448.

Churchill, B., & Craig, L. (2019). Gender in the gig economy: Men and women using digital platforms to secure work in Australia. *Journal of Sociology, 55*(4), 741-761.

Distelhorst, G., Hainmueller, J., & Locke, R. M. (2017). Does lean improve labor standards? Management and social performance in the Nike supply chain. *Management Science, 63*(3), 707-728.

Distelhorst, G., & McGahan, A. (2022). Socially irresponsible employment in emerging-market manufacturers. *Organization Science, 33*(6), 2135-2158.

Doorey, D. J. (2011). The transparent supply chain: From resistance to implementation at Nike and Levi-Strauss. *Journal of Business Ethics, 103*, 587-603.

Dubal, V. (2023). On algorithmic wage discrimination. *SSRN*. https://ssrn.com/abstract=4331080

Felstiner, A. (2011). Working the crowd: Employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment & Labor Law, 32*, 143-203.

Fieseler, C., Bucher, E., & Hoffmann, C. P. (2019). Unfairness by design? The perceived fairness of digital labor on crowdworking platforms. *Journal of Business Ethics, 156*, 987-1005.

Friedland, J., & Balkin, D. B. (2023). When gig workers become essential: Leveraging customer moral self-awareness beyond COVID-19. *Business Horizons*, 66(2), 181-190.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research, 23*(2), 135-162.

Graham, M., Woodcock, J., Heeks, R., Mungai, P., Van Belle, J.-P., du Toit, D., Fredman, S., Osiki, A., van der Spuy, A., & Silberman, S. M. (2020). The Fairwork Foundation: Strategies for improving platform work in a global context. *Geoforum, 112*, 100-103.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Horton, J. J., & Chilton, L. B. (2010). The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce* (pp. 209-218).

Karmarkar, U. S., & Apte, U. M. (2007). Operations management in the information economy: Information products, processes, and chains. *Journal of Operations Management, 25*(2), 438-453.

Kietzmann, J. H. (2017). Crowdsourcing: A revised definition and introduction to new research. *Business Horizons*, 60(2), 151-153.

Lehdonvirta, V. (2018). Flexibility in the gig economy: Managing time on three online piecework platforms. *New Technology, Work and Employment, 33*(1), 13-29.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition.

McGrath, P., McCarthy, L., Marshall, D., & Rehme, J. (2021). Tools and technologies of transparency in sustainable global supply chains. *California Management Review, 64*(1), 67-89.

McInnis, B., Cosley, D., Nam, C., & Leshed, G. (2016). Taking a HIT: Designing around rejection, mistrust, risk, and workers' experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2271-2282).

Merlo, O., Eisingerich, A., Auh, S., & Levstek, J. (2018). The benefits and implementation of performance transparency: The why and how of letting your customers 'see through'your business. *Business Horizons*, 61(1), 73-84.

Metters, R., & Verma, R. (2008). History of offshoring knowledge services. *Journal of Operations Management, 26*(2), 141-147.

Morreale, F., Bahmanteymouri, E., Burmester, B., Chen, A., & Thorp, M. (2023). The unwitting labourer: Extracting humanness in AI training. *AI & Society*, 1-11.

Morschheuser, B., & Hamari, J. (2019). The gamification of work: Lessons from crowdsourcing. *Journal of Management Inquiry, 28*(2), 145-148.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411-419.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1.

Prolific. (2024). Prolific's payment principles. Accessed February 20, 2024. https://researcher-help.prolific.co/hc/en-gb/articles/4407695146002

Prpić, J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. (2015). How to work a crowd: Developing crowd capital through crowdsourcing. *Business Horizons*, 58(1), 77-85.

Rani, U., & Furrer, M. (2021). Digital labour platforms and new forms of flexible work in developing countries: Algorithmic management of work and workers. *Competition & Change, 25*(2), 212-236.

Roberts, S. T. (2019). *Behind the screen*. Yale University Press.

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication, 10*, 27-44.

Schmidt, G. B., & Jettinghoff, W. M. (2016). Using Amazon Mechanical Turk and other compensated crowdsourcing sites. *Business Horizons*, 59(4), 391-400.

Schor, J. B., Attwood-Charles, W., Cansoy, M., Ladegaard, I., & Wengronowitz, R. (2020). Dependence and precarity in the platform economy. *Theory and Society, 49*, 833-861.

Shevchuk, A., & Strebkov, D. (2018). Safeguards against opportunism in freelance contracting on the Internet. *British Journal of Industrial Relations, 56*(2), 342-369.

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2023). The Curse of Recursion: Training on Generated Data Makes Models Forget. arXiv-2305.

Time. (2023, April 19). How an AI made its creators millionaires without any human input - But not without human exploitation. Accessed February 20, 2024. https://time.com/6247678/openai-chatgpt-kenya-workers/.

van Zoonen, W., Sivunen, A. E., & Treem, J. W. (2024). Algorithmic management of crowdworkers: Implications for workers' identity, belonging, and meaningfulness of work. *Computers in Human Behavior*, *152*, 108089.

van Zoonen, W., ter Hoeven, C., & Morgan, R. (2023). Finding meaning in crowdwork: An analysis of algorithmic management, work characteristics, and meaningfulness. *The Information Society*, *39*(5), 322-336.

Vaughan, J. W. (2018). Making better use of the crowd: How crowdsourcing can advance machine learning research. *Journal of Machine Learning Research*, *18*(193), 1-46.

Wang, J., Ipeirotis, P. G., & Provost, F. (2017). Cost-effective quality assurance in crowd labeling. *Information Systems Research*, *28*(1), 137-157.

Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, *61*(6), 823-832.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, *33*(1), 56-75.

Table 1. Risk ratios for perceiving MTurk as unfair for five metrics

Metric	"Worse" experience	% that find platform unfair		"Better" experience	Risk Ratio [95% CI]
Tasks engaging	Finding tasks "Very" or "Somewhat" boring	45.1%	7.5%	Finding tasks "Very" or "Somewhat" engaging	6.03* [4.35, 8.35]
Regret accepting tasks	>30% of accepted tasks	24.1%	11.2%	<30% of accepted tasks	2.15* [1.57, 2.94]
Time searching for tasks	>30% of time searching	19.5%	10.6%	<30% of time searching	1.85* [1.35, 2.54]
Obligation to accept tasks	"Often" or "Always"	18.6%	12.4%	"Never", "Rarely" or "Sometimes"	1.5* [1.09, 2.07]
Time for breaks	"About half the time", "Sometimes" or "Never"	15.4%	13.7%	"Most of the time" or "Always"	1.13 [0.82, 1.56]

Note: Risk ratios are calculated as the increase in likelihood that a respondent reports perceiving MTurk to be "somewhat unfair" or "very unfair" given they meet the specified criterion. For example, respondents who reported they found tasks on MTurk to be "somewhat boring" or "very boring" were six times as likely to find MTurk "somewhat unfair" or "very unfair" as those who did not. In the final column, an asterisk (*) indicates the risk ratio is statistically significant ($p<0.05$).

Figure 1. Despite obvious differences, the digital supply chain supporting AI displays some characteristics and challenges with parallels to physical supply chains

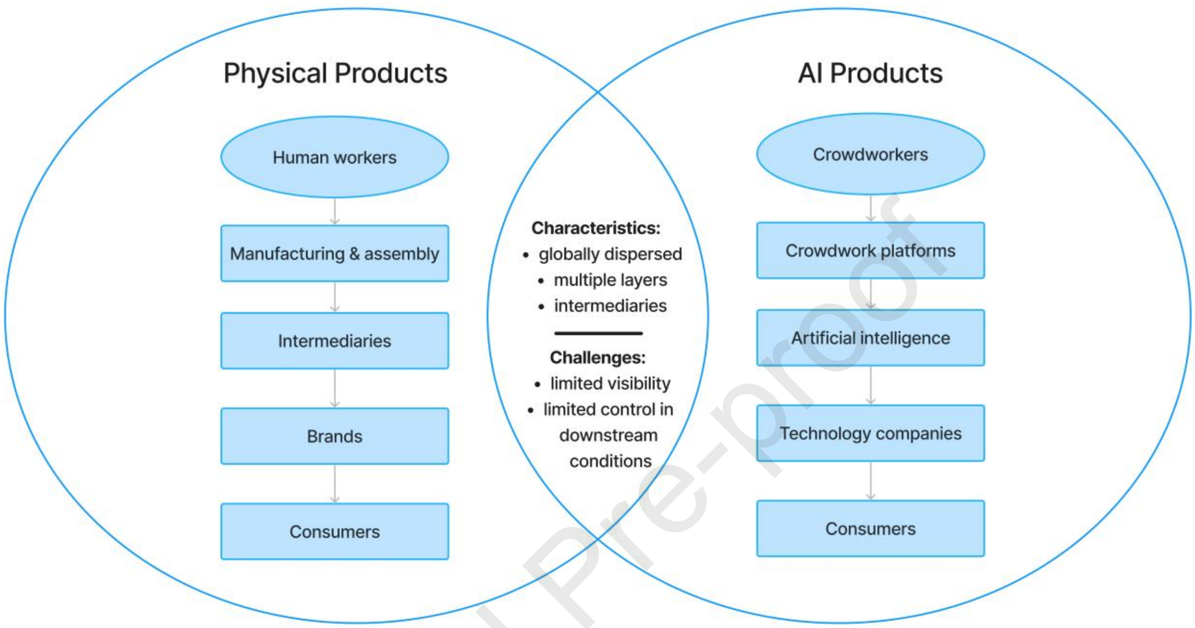
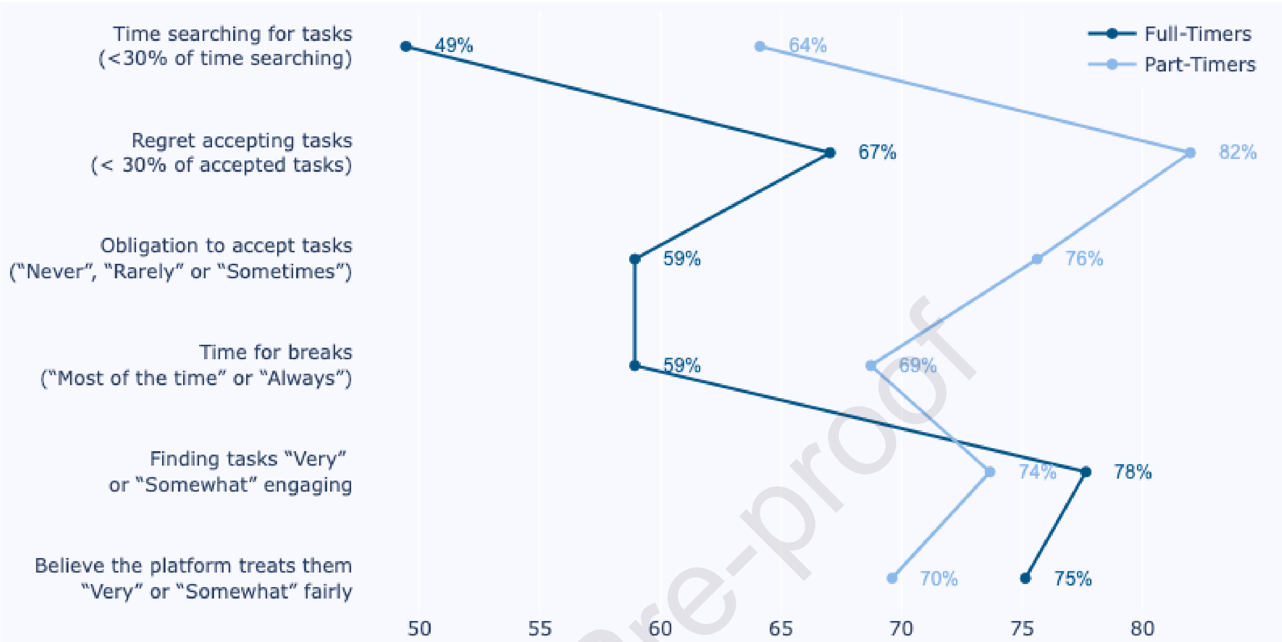
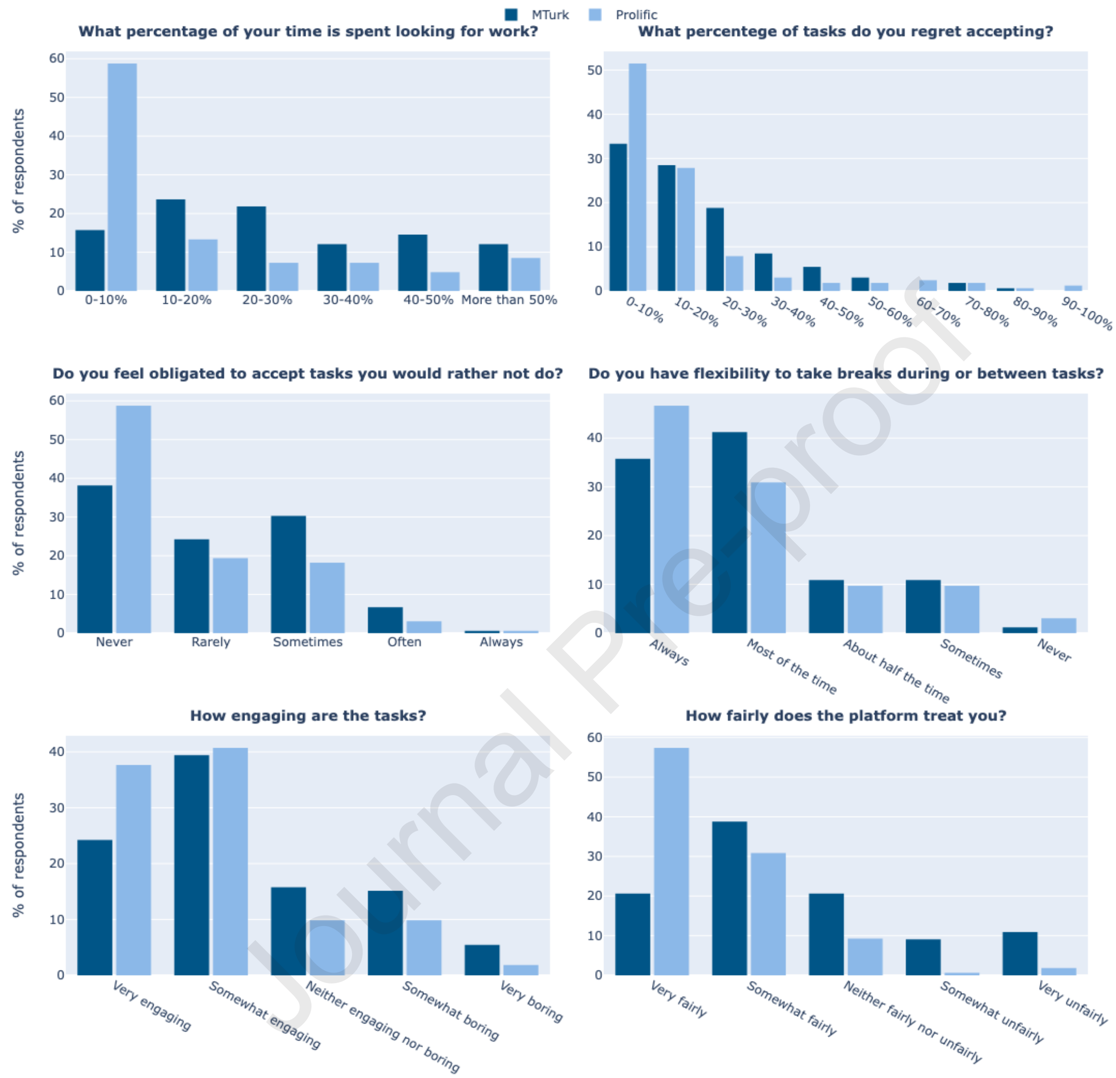


Figure 2. Full-time and part-time workers' experiences on MTurk



Note: The figure depicts the share of full-time and part-time workers who meet each of the six criteria listed. All differences are statistically significant ($p < 0.05$ using χ^2 test for independence of the distribution of Full-Timers and Part-Timers).

Figure 3. Comparison of MTurk and Prolific across six metrics



Note: Figure depicts the share of workers that selected each possible response for six main survey questions, for subset of n = 165 workers who reported working on both MTurk and Prolific. Difference in distributions are statistically significant ($p < 0.05$ using χ^2 test for independence of distributions for MTurk and Prolific responses) for all metrics except “flexibility to take breaks” ($p = 0.18$).