UNDERSTANDING WORKER PERCEPTION OF PAY FAIRNESS ON MICROTASK CROWDSOURCING PLATFORMS

YIDUO LIU

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN INFORMATION SYSTEMS AND TECHNOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

May 2026

© YIDUO LIU, 2026

Abstract

This thesis investigates the factors shaping worker perception of pay fairness on microtask crowdsourcing platforms. It employs a hybrid approach that integrates data-driven theory building with theory-driven sensemaking and empirical testing informed by organizational justice theory. Collecting and analyzing 14,553 reviews posted by Amazon Mechanical Turk workers on TurkerView, the study identifies seven latent themes each for positive and negative experiences and maps them onto distributive, procedural, and interactional justice dimensions. Distributive concerns (e.g., generous pay, bonus opportunities, easy earnings, and time efficiency) emerge as salient factors, while procedural and interactional factors reflect task design quality and requester communication. Predictive models, including stepwise econometric models and a DistilRoBERTa-based deep learning model, show that dense text embeddings capture fairness-relevant semantic information beyond topic-level abstractions in explaining pay fairness ratings. The findings advance understanding of pay fairness and inform platform design, requester practices, and labor policy in the gig economy.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor, Professor Ling Jiang, for her guidance, patience, and expertise throughout the course of this research. Her thoughtful feedback and consistent encouragement shaped both the direction and the quality of this work.

I am also grateful to Professor Xiaohui Yu for serving on my supervisory committee and providing constructive comments and valuable suggestions.

I wish to thank the Faculty of Graduate Studies and the Master's Program in Information Systems and Technology at York University for providing the academic environment and resources that made this research possible.

This study would not have been possible without the microtask workers who contribute reviews to TurkerView, whose candid reflections on their work experiences form the empirical foundation of this thesis. Their voices are central to this research, and I hope the findings contribute, however modestly, to improving fairness in online labor markets.

Finally, I would like to thank my family and friends for their unwavering support and understanding throughout my graduate studies.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background and Motivation

Microtask crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) have transformed labor markets by enabling global, on-demand workforces to perform granular, low-paying tasks (Wood et al., 2019). These platforms democratize access to employment by lowering geographic and skill barriers, offering labor opportunities to over 100,000 workers across 180 countries (*Amazon Mechanical Turk*, 2024). While microworkers benefit from the autonomy to choose tasks that fit their schedules (Deng et al., 2016), platform-based nonstandard work arrangements raise ethical concerns, such as unregulated wages and lack of worker rights protections, leaving workers vulnerable to exploitation (Tan et al., 2021).

Central to growing concerns about microwork is the debate over fairness in pay, driven by inconsistent earnings among workers. Recognizing that pay fairness can be perceived differently depending on social or deserved comparisons (Kim et al., 2019), this study focuses on worker perception of pay fairness within the context of microtask crowdsourcing platforms. Wage rates earned by MTurk workers, for instance, demonstrate a wide spectrum, ranging from a median wage of $1.38/hour (Horton & Chilton, 2010) to as high as $10.41/hour (Moss et al., 2023). The average hourly wage of microtasks across platforms is lower than $6 (Hornuf & Vrankar, 2022). In addition, invisible, unpaid labor (e.g., rejected tasks, searching for tasks, acquiring qualifications) further erodes actual wage rates (Fowler et al., 2023), yet workers' continued participation inadvertently perpetuates wage suppression, creating an ethical paradox. Some workers are willing to accept low-paying tasks due to non-monetary benefits such as flexibility and autonomy (Deng & Joshi, 2016; Jiang et al., 2015), which may influence their perception of pay fairness. Cultural norms, financial needs, and work experiences (Dornstein, 1989) further perplex worker perception of pay fairness, especially as the easy access to a global 24/7, on-demand workforce drives wages into a "race to the bottom" (Tan et al., 2021).

Existing research on pay fairness in microwork is constrained by conceptual and methodological limitations. Studies grounded in organizational justice theories often overlook the unique characteristics of platform-enabled nonstandard employment and the dynamics peculiar to online labor markets. Moreover, surveys used to solicit workers' perceptions are susceptible to biases inherent in self-reported data (Alpar & Osterbrink, 2020). The Fairwork Foundation (Graham & Woodcock, 2018) highlights systemic inequities at the platform level but lacks fine-grained insights into workers' individual evaluations.

1.2 Research Objectives

Motivated by the limitations in existing literature and the ethical paradoxes of the gig economy, this study addresses the overarching research question: *What factors affect worker perception of pay fairness on microtask crowdsourcing platforms?* Going beyond a single-phase analysis, this thesis pursues a comprehensive investigation through the following specific research questions:

- **RQ1:** *What latent factors emerge from crowdworkers' positive and negative experiences on microtask platforms?*
- **RQ2:** *How can the identified factors be aligned with the distributive, procedural, and interactional dimensions of organizational justice theory?*
- **RQ3:** *To what extent do the identified justice-related factors predict workers' perceived pay fairness?*
- **RQ4:** *How do econometric and deep learning models differ in their ability to predict workers' perceptions of pay fairness?*
- **RQ5:** *How do workers' perceptions of pay fairness on microtask platforms vary across workers with different levels of experience?*

These questions form the foundation of an integrated investigation into pay fairness in microtask crowdsourcing. They guide the study from inductively identifying salient themes in worker discourse, to interpreting those themes through the theoretical lens of organizational justice, and ultimately to

empirically evaluating their explanatory power through predictive modeling. In doing so, the study aims to develop a more nuanced understanding of pay fairness that is both theoretically grounded and empirically supported.

1.3 Research Method

To achieve these objectives, this study employs a hybrid approach that integrates data-driven, computationally intensive theory building (Berente et al., 2019) with theory-driven sensemaking and empirical testing informed by organizational justice theory. The research design is grounded in an empirical analysis of microworker reviews collected from TurkerView, a third-party review platform designed specifically for the MTurk community. This data source provides a rich, unprompted repository of worker sentiment that circumvents the biases often found in traditional solicited surveys. The hybrid approach comprises two integrated components:

- **Data-driven theory building:** Using computationally intensive topic modeling (StableLDA) to inductively identify latent factors from worker-generated text, followed by a systematic, theory-driven mapping of these factors onto the dimensions of organizational justice.
- **Theory-driven empirical testing:** Evaluating the predictive power of the identified justice-related factors through econometric models (stepwise OLS regression) and deep learning models (transformer-based language model architectures), enabling a direct comparison of interpretable and high-performance approaches.

Aligned with the hybrid approach, the data analysis is organized into two major phases:

- **Phase 1 – Factor Identification via Topic Modeling**: StableLDA is applied to extract latent themes from workers' positive and negative reviews, and the resulting topics are mapped onto the distributive, procedural, and interactional dimensions of organizational justice.
- **Phase 2 – Predictive Modeling of Pay Fairness**: Econometric models (pooled OLS with topic probabilities) and deep learning models (DistilRoBERTa-based DNN with text embeddings) are

developed to predict workers' numerical pay fairness ratings, with pace-stratified analyses to examine heterogeneity across experience levels.

The analytical process utilizes topic modeling to extract key themes from these reviews, allowing for the inductive identification of factors that define the microwork experience. Building on the psychometric measure developed in prior organizational justice research (Colquitt, 2001), this study systematically maps the factors emerging from topic modeling onto established dimensions of organizational justice. This enables an alternative operationalization of the fairness construct, customized for microtask crowdsourcing. By combining the inductive discovery of themes emerging from the data with the deductive evaluation of their predictive power, this hybrid approach offers a robust framework for understanding the determinants of worker perception of pay fairness in microtask crowdsourcing.

1.4 Research Contributions

This study offers several key contributions to both theory and practice:

- **Theoretical extension of organizational justice to labor in microtask platforms.** This research applies Colquitt's (2001) three-factor justice model to microwork, demonstrating that distributive, procedural, and interactional justice dimensions remain salient even in algorithmically mediated gig environments where traditional employer–employee relationships are absent. At the same time, it identifies justice-relevant factors that are distinctive to the microwork context, thereby extending the conventional understanding of organizational justice theory and illustrating how justice perceptions are reinterpreted in platform-based labor settings.
- **A hybrid methodology that balances interpretability and predictive power.** By integrating StableLDA topic modeling with pooled OLS econometrics and a DistilRoBERTa-based deep learning architecture, this study demonstrates the complementary value of the two approaches. The topic-model-based econometric analysis offers theoretically interpretable evidence regarding the relationship between justice-related factors and workers' pay fairness ratings, whereas the deep

learning model provides greater predictive power by capturing semantic patterns that exceed topic-level abstractions. In this way, the study shows how unsupervised text analysis can be effectively paired with established social science frameworks to extract and quantify fairness-relevant signals from unstructured worker-generated text. Taken together, these approaches combine explanatory interpretability with predictive strength in the analysis of workers' fairness perceptions.

- **Actionable insights for platform design, requester practices, and labor policy.** By uncovering the factors that enhance or erode perceived pay fairness, this study offers practical implications for improving the microwork environment. The findings show that fairness perceptions are influenced not only by compensation levels, but also nonmonetary factors such as excessive task length, interface errors, and inadequate requester communication. These insights provide an empirical basis for platform operators to refine platform design and oversight mechanisms, for requesters to adopt fairer and more transparent task management practices, and for policymakers to consider stronger standards for compensation fairness and worker protection in digitally mediated labor markets.

1.5 Thesis Outline

The remainder of this thesis is organized as follows.

Chapter 2: Literature Review reviews the relevant literature underpinning this study. It begins by examining the characteristics and pay dynamics of microtask crowdsourcing platforms, followed by a discussion of workers' evaluations of pay. The chapter then provides a detailed account of organizational justice theory and its three dimensions, namely distributive, procedural, and interactional justice, as they apply to platform-mediated work. It concludes with reviews of topic modeling in information systems research and deep learning models for text analysis.

Chapter 3: Methodology describes the research design and analytical procedures. It details the data collection process from TurkerView, including the web scraping and preprocessing steps. The chapter then

presents the two phases of data analysis: Phase 1 (factor identification via StableLDA topic modeling) and Phase 2 (predictive modeling using pooled OLS regression and a deep learning architecture based on DistilRoBERTa embeddings).

Chapter 4: Topic Modeling Results presents the findings from Phase 1 of the analysis. It reports the optimal number of topics derived from five evaluation metrics and details the seven-topic solutions for both positive and negative worker experiences. Each topic is interpreted with representative quotes, and the chapter concludes with the theory-driven alignment of these topics to the dimensions of organizational justice.

Chapter 5: Predictive Modeling Results presents the findings from Phase 2. It reports the pooled and pace-stratified OLS regression results, followed by the performance comparison of deep learning models. The chapter demonstrates the substantial improvement in explanatory power achieved by transformer-based text embeddings over topic probability features.

Chapter 6: Discussion, Contributions and Implications interprets the findings from both phases of analysis, presents a summary of key findings, and articulates the theoretical contributions and practical implications of the research for platform designers, requesters, and policymakers.

Chapter 7: Conclusions, Limitations and Future Work concludes the thesis by summarizing the overall argument, acknowledging the limitations of the study, and proposing directions for future research, including cross-platform validation, incorporation of negative experience factors, and exploration of the interplay between perceived and normative fairness.

Chapter 2. Literature Review

2.1 Microtask Crowdsourcing Platforms

2.1.1 Characteristics of Microtask Crowdsourcing Platforms

Microtask crowdsourcing platforms facilitate the distribution of granular, independent units of work, known as microtasks, to a distributed global workforce. These platforms function as two-sided marketplaces connecting requesters, who break down complex projects into manageable components such as image annotation, survey completion, or data verification, with workers who complete these tasks for monetary compensation (Difallah et al., 2018). The workforce is notably diverse, spanning various demographics and geographies, often comprising educated individuals seeking flexible supplemental income or a primary livelihood in regions with limited local employment (Berg et al., 2018). Among these platforms, Amazon Mechanical Turk (MTurk) serves as a seminal example, operating on an open call model where workers browse and select tasks at will.

However, the nature of this work environment is characterized by a distinct duality (Deng et al., 2016). While providing open access to labor markets and empowering workers with flexibility and autonomy over their schedules, microtask crowdsourcing platforms give rise to significant ethical concerns. The lack of formal labor protections and collective bargaining power often leaves workers vulnerable to structural injustice, creating a persistent tension between the platform's empowering potential and its marginalizing effects (Chiu et al., 2022; Deng et al., 2016).

2.1.2 Pay Levels on Microtask Crowdsourcing Platforms

The architecture of microtask crowdsourcing platforms represents a shift from traditional time-based employment to a fine-grained piece-rate labor model. On crowdsourcing platforms like MTurk, Prolific, and emerging AI-training ecosystems, the determination of wages is seemingly a function of market dynamics. Clients who post tasks, or "requesters", can set pay rates at their discretion for individual tasks

(Human Intelligence Tasks, or HITs), creating a spot market for labor where prices fluctuate based on demand and task complexity (D. L. Chen & Horton, 2016). However, this market is characterized by significant asymmetries. While requesters set nominal rates, workers exhibit highly heterogeneous earning capabilities, driven by variations in skill, experience, internet speed, and the use of auxiliary software.

Piece-rate payment is the prevalent compensation model in microtask crowdsourcing platforms, where workers are paid a fixed sum per completed unit rather than for the time invested. Empirical analyses of large-scale platform data indicate that these rates frequently average between $0.10-$0.20 per task (Hara et al., 2018; Marshall et al., 2023). This further complicates the evaluation of task-based earnings in terms of comparable time-based rates. The interaction between posted piece-rates, supply elasticities, and workers' reservation wages produces a wide dispersion in realized pay, even within the same task class (J. Yang et al., 2018). Foundational research by Horton and Chilton (2010) demonstrates that workers in these markets behave according to rational labor supply models, estimating a geometric mean reservation wage of $1.49/hour in lower-difficulty tasks. However, the structural design of the platform encourages a "race to the bottom" in pricing, where workers' willingness to accept tasks at depressed rates undermines collective wage standards.

This granularity in pay level has created a complex relationship between financial incentives and worker performance in microtask crowdsourcing. Laboratory and field experiments have extensively probed whether higher pay correlates with better work. The consensus in the literature indicates that while higher pay consistently increases the quantity of labor supplied (effort) and the speed at which tasks are accepted (uptake), its effect on quality (accuracy) is not significant (Cobanoglu et al., 2021). Instead of financial incentives, accuracy in crowdsourcing tasks often depends on task difficulty, the clarity of instructions, and the intrinsic meaningfulness of the work itself (Chandler & Kapelner, 2013; Mason & Watts, 2010). This creates a perverse incentive structure: workers who prioritize speed often achieve higher hourly returns than those who are meticulous, provided their work meets the minimum threshold for acceptance (Di Gangi et al., 2024). Interestingly, expert workers often develop heuristics and scripts that allow them to be both faster

and more accurate than novices, a phenomenon that reinforces inequality through the mechanism of piece-rate payment (Dupuis et al., 2022). This interaction between pay, intrinsic motivation, and heterogeneous worker speed helps explain why nominally "fair" piece-rates can yield inequitable time-based wages across workers.

The divergence of posted and realized wages is also troubling to workers. A critical turning point in the understanding of crowdwork economics was the data-driven analysis by Hara et al. (2018), which utilized a browser plugin to track the actual workflow of 2,676 workers performing over 3.8 million tasks. This study exposed a stark bifurcation between the wages requesters believe they are paying and the wages workers actually receive. This analysis revealed that while the average requester pays approximately $11.58 per hour based on task completion times, the median worker earns only ~$2.00 per hour. This dramatic discrepancy is driven by the "long tail" of low-paying requesters who dominate the platform's volume. Although a minority of high-paying requesters exist, the vast majority of available tasks offer low wages. Specifically, the study found that only 4% of workers earned more than the U.S. federal minimum wage of $7.25 per hour. This finding challenges the narrative that low wages are solely a function of unskilled labor; rather, it points to structural inefficiencies where high-volume, low-wage tasks drown out fair opportunities.

A separate meta-analysis of 42 studies synthesizing crowdsourcing wage data found an average hourly wage of $4.97 among microtask workers, dropping to $3.78 if unpaid work is accounted for, with substantial variability across platforms and task types (Hornuf & Vrankar, 2022). This wage correction factor proved essential for accurate research, as earlier studies that ignored invisible labor systematically overestimated worker well-being.

On the other hand, much of workers' time is spent on unpaid work, or "invisible labor". In traditional employment, activities such as searching for tasks, communicating with supervisors, and managing administrative overhead are compensated components of the workday. In the gig economy, these are externalized onto the worker. Approximately 38% of workers report spending over half their work time on unpaid activities like searching for suitable work or communicating with requesters, reducing realized

wages by 30-40% (Toxtli et al., 2021). A stream of empirical studies quantifies realized hourly earnings after accounting for unpaid time, supporting concerns that nominal piece-rates mask substantial effective wage suppression: Hara et al. (2018) analyze large-scale MTurk traces and estimate median hourly earnings near or below US minimum wage once task searching and qualification time are included, aligning with evidence from the International Labour Organization's cross-platform audit showing that extensive unpaid time depresses realized wages across regions and task types (Berg et al., 2018).

2.2 Workers' Evaluations of Pay

Alongside the investigations of wage rates, prior research on microtask crowdsourcing has also examined workers' perceived fairness in pay, which is defined as workers' evaluations of whether the compensation aligns with invested time and effort (Salminen et al., 2023). While economic rationality provides a baseline for understanding labor supply, it fails to fully explain the persistent workforce attendance with the sub-minimum wages described in the previous section. This line of research has examined the impacts of worker perception of pay fairness on outcomes, such as performance, satisfaction, and retention in crowdwork (Hara et al., 2018; Toxtli et al., 2021), and identified its antecedents, including workers' expectations and reference-point comparisons contexts (Kim et al., 2019).

2.2.1 Antecedents of Fairness Perception

The formation of fairness perceptions in crowdwork is a complex cognitive process influenced by both structural and psychological factors. Unlike traditional employees who may compare salaries with colleagues (social comparison) (Festinger, 1954), crowdworkers operate in relative isolation (Vallas & Schor, 2020). Consequently, their reference points for fairness are often internal or abstract.

Recent studies suggest that in the absence of visible peers, workers rely on three major aspects: Self-reference, which is the comparison between current task earnings with workers' own historical average or daily earnings goal (Camerer et al., 1997; M. K. Chen, 2016); Algorithmic signals, where gamified elements like leaderboards or "percentile" feedback can induce social comparison, often leading to unflattering

comparisons that reduce satisfaction for the majority of workers who naturally fall below the top tier (Behl et al., 2021; Gandini, 2019); And institutional anchors, when workers often use the minimum wage of their home country or the "reservation wage" derived from their alternative employment options as a benchmark (D'Cruz & Noronha, 2016; Wood et al., 2018).

Aside from these reference points, new research by Scheel et al. (2025) highlights that fairness perceptions are not solely about the money but about the experience of earning it. Utilizing the Job Characteristics Model (JCM), they found that the mitigation of amotivation (active disengagement or frustration) is more predictive of platform commitment than traditional motivators like autonomy. When tasks are poorly designed, repetitive, or prone to technical errors, workers perceive the pay as less fair because the "psychological cost" of the labor is higher. Conversely, well-designed tasks can buffer the negative impact of low wages.

2.2.2 The Reframing of "Work vs. Leisure"

A significant theoretical advancement in understanding how workers reconcile low pay is provided by Jiang and Wagner (2024). Jiang and Wagner address the paradox of why highly educated workers (often with degrees) participate in microwork for pennies. Their findings suggest that workers often engage in a cognitive reframing of the activity. First, workers conceptualize microtasking as occupying an ambiguous space between "work" and "leisure". It is often performed during downtime, while watching television, or in between other activities. Next, the study introduces the concept that workers assign a low opportunity cost to this specific block of time. If the alternative activity is "unproductive leisure" (e.g., scrolling social media) which pays $0, then a microtask paying $2.00/h is perceived as a relative gain. Finally, by framing the activity as "productive leisure", workers lower their reference point for fairness. They are not comparing the $2.00/h to their professional salary ($30/h), but to the $0/h of leisure. This blurring of boundaries allows workers to maintain satisfaction despite objective economic exploitation, which influences how workers perceive pay fairness in crowdsourcing.

2.2.3 Consequences of Perceived Unfairness

Empirical studies further demonstrate that piece-rate compensation can yield inequality in hourly wages due to variation in task completion times across workers (Saito et al., 2019; Salminen et al., 2023), thereby reinforcing workers' perceptions of unfairness even when nominal rates appear adequate (Hornuf & Vrankar, 2022).

This duality has brought complex consequences over the years. To counter low pay rates, workers adapt through strategies. Third-party transparency tools, like TurkerView and Turkopticon, are developed to help workers identify worthwhile work and avoid abusive requesters. These tools partially mitigate information asymmetries by creating a reputation system for employers, allowing workers to filter out "wage thieves" (requesters who reject work to avoid payment) (Hornuf & Vrankar, 2022; Savage et al., 2020). In cases of perceived gross injustice (e.g., unexplained rejections), workers may resort to retaliatory low-quality submissions, or "spamming" (Barends & de Vries, 2019; Hara et al., 2018), which highlights the tension between reducing labor costs and maintaining work quality on microtask crowdsourcing platforms.

However, structural power remains with requesters who control acceptance, rejections, and pay (Irani & Silberman, 2013, 2016). While worker-led reputation infrastructures provide a safety net, they cannot eliminate wage risks that are tied to rejection and unexpected requester behavior. These tools' partial efficacy underscores how platform-level governance, rather than individual coping strategies alone, shapes distributional outcomes in microwork marketplaces (Berg et al., 2018). Although scholars emphasize the importance of aligning compensation with ethical standards (Kummerfeld, 2021; Whiting et al., 2019), platform policies remain stagnant, lacking meaningful systemic reforms. This is evident in the absence of payment regulation on MTurk (Pittman & Sheehan, 2016) and the continued classification of workers as independent contractors to avoid minimum wage laws, underscoring the need for a deeper understanding of worker perception of pay fairness.

Why do workers continue to engage, and how do they construct a sense of justice in an unjust system? To explore potential factors influencing worker perception of pay fairness, this study draws upon organizational justice theory as its primary theoretical framework, reviewed in the next section.

2.3 Organizational Justice Theory

Developed from research in traditional workplaces, organizational justice theory offers a robust, multidimensional framework for understanding how people judge the fairness of workplace events and interactions (Greenberg, 1990). Applying this theory to the growing gig economy and online labor platforms is an active and important area of recent research, providing key tools to analyze the unique fairness challenges found in new forms of work managed through online platforms. This section reviews the basic principles of organizational justice theory and brings together recent studies that use this framework in online work environments, showing why it is a good fit for analyzing fairness perceptions in microtask crowdsourcing.

Organizational justice generally means how an individual sees the fairness and ethics of how managers behave and how an organization operates (Cropanzano et al., 2007). While the words *justice* and *fairness* are often used interchangeably, it is helpful to distinguish between them: Justice typically refers to adherence to established rules and the equitable distribution of resources, rewards, and punishments. In contrast, fairness pertains to impartiality and the equitable treatment of individuals, taking into account specific contexts and circumstances (Colquitt, 2001). Fairness is generally considered more subjective, focusing on the perceptions and lived experiences of individuals. This study is mainly concerned with the latter: How workers perceive the fairness of microwork.

Organizational justice theory identifies three dimensions of justice, namely distributive, procedural, and interactional justice (Greenberg, 1990), as a framework for analyzing microworker experiences. Prior research has shown that fairness perceptions are not a single, simple construct, but a multidimensional construct shown by the three-part model by Colquitt (2001). This model is built on the foundational work

including equity theory, which posits that individuals evaluate outcome fairness through input–output ratios relative to referents (Adams, 1965), and fair procedures, which are directly testable criteria in platform governance (consistency, bias suppression, accuracy, correctability, representativeness, and ethicality) (Leventhal, 1980). Norms of respectful treatment and adequate explanations during decision processes also affect the perception of fairness later formalized by Bies (1986). These concepts are all itemized in Colquitt's (2001) multidimensional measure. Each dimension is examined in turn below.

2.3.1 Distributive Justice

Distributive justice focuses on the extent to which outcomes are allocated equitably based on effort, contribution, or performance. This dimension is theoretically grounded in equity theory, which posits that individuals evaluate outcome fairness by creating a ratio of their inputs to their outputs and then comparing this ratio to that of a referent (Adams, 1965). Inputs are what the individual invests, such as effort, time, skill, and experience. Outputs are what they receive in return, such as pay, recognition, and status. According to Adams (1965), a state of inequity is perceived when an individual's input/output ratio is unequal to that of their comparison other, leading to psychological distress and motivating actions to restore balance.

This equity-based framework is directly applicable to microwork, though the components of the input/output ratio are uniquely defined by the platform context. For a microworker, "inputs" are not just the billable time-on-task. They also include a significant amount of unpaid "overhead" labor, such as the time spent searching for tasks, reading complex instructions, learning new platform interfaces, and managing rejections (W.-C. Chen et al., 2019). The "outputs" are primarily the low pay for microwork, but also include non-monetary outcomes such as reputation scores, access to future tasks, and, occasionally, feedback. This broader conceptualization of inputs and outputs is essential for accurately capturing the worker's "realized hourly pay" (Kittur et al., 2013).

The comparison referent is also fundamentally destabilized in the anonymous, globalized, and de-contextualized microwork market. In a traditional office, the referent is typically a co-worker in a similar role. For a microworker, the referent is ambiguous. Is it another anonymous worker on the platform, who may be in a different country with a different cost of living? Is it the worker's own past earnings? Is it the legal minimum wage in the worker's own jurisdiction, which platforms often ignore? Or is it a conceptual "fair wage" for the task itself? This ambiguity and lack of a stable referent means that workers may grasp for any available standard, leading to high variability and volatility in fairness perceptions.

Violations of distributive justice are pervasive and a central grievance on microtask crowdsourcing platforms. The most salient issue is the low pay for microwork, which often falls well below minimum wage standards in the workers' home countries (Kittur et al., 2013). This low pay creates a fundamental perceived misalignment between the high inputs (effort, time, overhead) and the low outputs (pay). This perception of distributive injustice is not passive; it directly influences worker behavior, undermining their subsequent engagement with microtasks and their willingness to remain on the platform (Alpar & Osterbrink, 2020).

This misalignment is severely exacerbated by the prevalence of unpaid labor, particularly in the form of rejected tasks. Post-task evaluations from crowdworkers reveal declining satisfaction, largely due to this unpaid labor (Alpar & Osterbrink, 2020). A task rejection represents a total failure of distributive justice: the worker has provided 100% of the input (time, effort, skill) for 0% of the output (pay). This creates an infinitely inequitable ratio and highlights a critical feature of microwork: distributive justice is not merely low, it is contingent and uncertain. In traditional employment, an employer purchases a worker's time (input) and bears the economic risk of that time being unproductive. In microwork, the requester purchases only accepted output, effectively transferring 100% of the economic risk for ambiguous instructions, technical failures, or subjective evaluation onto the worker. This shifting of risk is a profound form of distributive injustice, existing even before the low pay rate is considered.

2.3.2 Procedural Justice

Procedural justice concerns the perceived fairness of the processes used to determine outcomes, emphasizing elements such as voice, consistency, correctability, and ethicality in decision-making. This dimension is theoretically grounded in the work of Leventhal (1980), who articulated six key criteria that individuals use to judge the fairness of a procedure: consistency (applied uniformly across people and time), bias suppression (neutral and objective), accuracy (based on good information), correctability (mechanisms to appeal or correct bad decisions), representativeness (giving "voice" to participants affected by the decision), and ethicality (adherence to moral and ethical norms). Procedural justice theory posits that individuals are more willing to accept a negative outcome (low distributive justice) if they believe the process used to reach that outcome was fair.

In the context of online crowd work, these concepts are directly testable in platform governance, as the platform itself, through its design and rules, becomes a main source of fairness perceptions, especially in the areas of process and information (Lee et al., 2023). Unlike traditional workplaces where procedures are social policies enacted by human managers, in microwork, the platform's code, algorithms, and automated systems are the procedures. This insight moves the study of procedural justice from the purely organizational and managerial realm into the socio-technical realm of human-computer interaction (HCI) and platform design.

Because the platform is the procedure, violations of procedural justice are often systemic and built directly into the platform's architecture. Procedural justice is compromised by "algorithmic management" (Duggan et al., 2020), which facilitates task assignment and evaluation while often operating as a black box. Workers are subjected to algorithmic decisions that determine their reputation, access to work, and pay, with no ability to assess the consistency, bias suppression, or accuracy of these systems (Leventhal, 1980). This opacity makes it impossible for workers to trust the system.

The most significant and pervasive violation of procedural justice on microwork platforms is the lack of correctability. Platforms are defined by their limited appeal mechanisms (Schulze et al., 2023). When a task is rejected, which is a decision that directly impacts the worker's distributive justice, there is often no formal, transparent, or timely process to appeal the decision. This lack of recourse is a direct violation of Leventhal's (1980) criterion and also suppresses representativeness, as it denies workers a voice in the decisions that most affect them.

The perception of procedural justice has a powerful, direct effect on worker behavior, acting as a gateway for worker engagement. In a high-uncertainty environment like microwork, workers lack complete information to know if a requester will ultimately provide a fair outcome (distributive justice). They therefore use the visible elements of the work, such as the clarity of instructions, the transparency of rules, and the perceived fairness of the platform's processes, as a heuristic to predict the trustworthiness of the requester and the likelihood of future distributive justice. Research in the online crowd work context has shown that high procedural justice (e.g., transparent rules) enhances pre-task fairness perceptions, which can increase task acceptance rates by 1.5 times (Alpar & Osterbrink, 2020).

2.3.3 Interactional Justice

Interactional justice refers to the perceived fairness of the treatment individuals receive in the workplace during the enactment of procedures and the communication of outcomes. This concept, first formalized by Bies (1986), highlights that the social and communicative aspects of an interaction can be as important as the procedures and outcomes themselves. Later research (Colquitt, 2001) divided interactional justice into two distinct sub-dimensions: interpersonal justice and informational justice, which are both itemized in Colquitt's multidimensional measure.

Interpersonal justice addresses the quality of interpersonal interactions and the extent to which individuals are treated with respect, dignity, and politeness. In the microwork context, where communication is often anonymous, asynchronous, and text-based, this dimension is frequently violated. Interactional justice is

impaired by dismissive communication from requesters, who hold a structural position of power (Fowler et al., 2023). The platform environment can disinhibit an impaired and dismissive communication style that would be socially unacceptable in a traditional, face-to-face workplace. Worker-led platforms and forums are replete with examples of workers being treated as disposable, replaceable, or non-human, a clear failure of interpersonal justice (Irani & Silberman, 2013).

Informational justice relates to the perceived fairness of communication, specifically the candidness, adequacy, and timeliness of explanations for decisions. This is a critical point of failure in microwork and is inextricably linked to the procedural and distributive justice failures previously discussed. A common informational justice violation is requesters' "refusal to provide feedback or explain the reasons for rejected work" (Fowler et al., 2023). This lack of prompt, and informative requester communications (Kittur et al., 2013) leaves workers in a state of ambiguity, unable to understand why their labor was devalued, how to improve, or whether the rejection was arbitrary. This practice is not only deeply frustrating but actively erodes trust, reinforces existing power imbalances, and threatens the long-term sustainability of crowdsourced labor ecosystems (Fowler et al., 2023).

Informational justice plays a unique role as a high-leverage buffer that can, when present, mitigate the negative effects of procedural and distributive injustice. A rejected task is a failure of distributive justice (no pay) and often a failure of procedural justice (no appeal). In the absence of an explanation (low informational justice), a worker is likely to attribute this failure to a malicious requester or a biased system, leading to profound distrust. However, if the requester provides high informational justice, that is, a candid, specific, and respectful explanation for the rejection, it can reframe the worker's perception. The worker may still be unhappy with the distributive outcome (no pay), but the explanation may lead them to see the procedure as accurate (Leventhal, 1980) (e.g., "I did make a mistake"). This prevents the worker from assuming the worst and helps preserve the relationship, demonstrating the powerful, trust-building function of simple explanations.

The systemic failure of platforms and requesters to provide interactional justice has spurred the creation of worker-led interventions, a form of collective action to re-institutionalize justice from the bottom up. Reputation and reviewing platforms, such as TurkerView, function as a parallel system to fill this void (Irani & Silberman, 2013). These platforms bolster interactional justice by incentivizing civility and explanations, as requesters who wish to maintain a good reputation (and thus attract good workers) are forced to be more respectful and communicative. These systems also indirectly improve procedural justice by deterring arbitrary rejections. This demonstrates that workers are not passive recipients of injustice but are actively building their own infrastructure to create accountability and enforce justice norms in an unregulated market.

2.3.4 A Three-Factor Justice Model for Microwork

Building on the established organizational justice literature, this study adopts Colquitt's (2001) three-factor framework to analyze the microwork experience. While the preceding discussion outlined the theoretical origins of justice perceptions, the utility of Colquitt's model for this study lies in its ability to provide construct clarity and measurement guidance for mapping crowd-work experiences to fairness perceptions. Colquitt et al. (2013) further show that justice dimensions are correlated but non-redundant and differentially predict attitudes and behaviors such as satisfaction, commitment, and rule compliance, justifying separate operationalization and testing in novel contexts like microwork. This evidence base supports the study's strategy to map microworker experiences to distinct justice dimensions while allowing for cross-dimension interactions.

Previous studies have also identified nuanced effects and intricate interactions among different dimensions of justice. In traditional workplaces, while distributive justice acts as a baseline expectation, procedural justice plays a significant role in predicting worker retention through transparent decision-making processes (Skarlicki & Folger, 1997). Some employees even prioritize interactional justice (e.g., respectful treatment) and procedural justice (e.g., appeal routes) over distributive justice (e.g., justified wage) (Beugre & Baron,

2001). In the context of online crowd work, procedural justice enhances pre-task fairness perceptions, increasing task acceptance rates by 1.5 times, whereas post-task evaluations reveal declining satisfaction, largely due to unpaid labor such as rejected tasks (Alpar & Osterbrink, 2020), highlighting concerns about distributive justice. Fieseler et al. (2019) further show that platform design can affect distributive, procedural, and interactional justice perceptions. Findings from prior research on justice in various work settings indicate that justice dimensions function as an interdependent system rather than as isolated components. Therefore, to understand pay fairness, it is necessary to look beyond the pay rate and consider the procedural and interactional context surrounding that pay. Table 1 synthesizes representative organizational justice research in online labor contexts, summarizing the justice dimensions investigated, the platform features and antecedents examined, the key worker outcomes documented, and the relevance of each study to microwork pay fairness.

Author(s) & Year	Research Context	Justice Dimensions Investigated	Key Antecedents & Platform Features	Key Outcomes & Worker Responses	Relevance to Microwork Pay Fairness
Song et al. (2020)	Crowd-based labor (general)	Distributive, Procedural, Interactional	Compensation policy, Participative & rule-based evaluation, Interactive & considerate communication	Requester turnover, Platform turnover	Provides a comprehensive conceptual model linking specific platform policies and communication styles directly to the three core justice dimensions and worker retention.
Jabagi et al. (2025)	Ridesharing	Interpersonal, Distributive, Informational	Customer disrespect (interpersonal injustice), Low customer rating (distributive injustice)	Negative emotions (anger, unhappiness), Impaired driving & service performance (real-time and carry-over effects)	Demonstrates that justice violations (especially interpersonal) have immediate emotional and behavioral consequences that can persist across tasks, highlighting the importance of requester-worker interactions.
Ye et al. (2017)	Microtask (MTurk)	Perceived Fairness in Pay (PFP)	Payment amount (high vs. low)	Performance quality (for effort-based tasks), Satisfaction, Task completion time	Establishes PFP as a critical mediator between pay rate and worker outcomes, showing that higher pay improves performance and satisfaction by making workers feel fairly treated.
Alpar & Osterbrink (2020)	Microtask (MTurk)	Perceived Fairness in Pay (PFP)	Crowdworker education, Crowdworker experience, Prior experience with requester	Task acceptance, Post-task fairness perceptions	Identifies worker-specific characteristics and relational history as key antecedents of pre-task fairness perceptions, influencing whether a worker even accepts it.
Deng et al. (2016)	Microtask (MTurk)	Fairness (as a core value)	Compensation structures, Platform governance, Technology design, Microtask characteristics	Empowerment vs. Marginalization	Identifies fairness as one of nine core values for microworkers. Shows that platform design features related to compensation and governance directly impact whether workers feel empowered or marginalized.

Table 1. Synthesis of Organizational Justice Research in Online Labor Contexts

Platform features and requester behaviors can be systematically linked to justice dimensions to guide measurement and design interventions. Distributive justice in microwork is reflected in alignment between time-on-task (including unpaid overhead) and realized hourly pay, procedural justice in transparent, consistent, and correctable allocation and evaluation processes, and interactional justice in respectful, prompt, and informative requester communications (W.-C. Chen et al., 2019; Kellogg et al., 2020; Kittur et al., 2013). Worker-led reviewing platforms such as TurkerView bolster interactional justice by incentivizing civility and explanations, while also indirectly improving procedural justice by deterring arbitrary rejections (Irani & Silberman, 2013, 2016).

2.4 Topic Modeling in Information Systems

In light of the hybrid approach, this research uses topic modeling to find factors directly from worker experiences and then connect them to justice theory. Topic modeling has become a typical computational method for inductive theory building from large, unstructured text corpora, enabling researchers to extract emergent themes from user-generated content, reviews, and platforms at scale (Berente et al., 2019; Hannigan et al., 2019). In IS and adjacent management research, topic models are used to discover concerns, practices, and sentiments in textual data that would be prohibitively costly to code manually, thereby informing construct development and explanatory models (Hannigan et al., 2019). This section reviews the basic ideas of topic modeling and its use as a tool for developing theory, explaining why it was chosen as the main method for finding key themes in microworker reviews.

Topic modeling refers to a group of automatic computational methods designed to analyze a collection of documents and find the abstract "topics" within them, with Latent Dirichlet Allocation (LDA) being the most fundamental (Blei et al., 2002). The main idea behind LDA is that documents are made of a mix of hidden topics, and each topic is a set of words, each with a certain probability. By analyzing how often words appear together, LDA can figure out which words tend to be grouped, clustering them into topics

that a human researcher can then interpret and name (Gillies et al., 2022). Because it requires human input and interpretation, topic modeling is a powerful tool that helps rather than replaces qualitative analysis.

The use of topic modeling in IS and management research has grown from simple content summaries to become an advanced method for building and expanding theory from text, often described as a data-driven, computationally-intensive theory development process (Berente et al., 2019). This approach has been more formally named Computational Grounded Theory, a method that combines computational tools like topic modeling with interpretive ideas of traditional grounded theory (Nelson, 2020; Odacioglu & Zhang, 2022). It usually involves a multi-step process of finding patterns with algorithms, refining them through close reading, and confirming them with more computational analysis, using the best parts of both machine and human analysis.

To further illustrate the practical application of these methods, Table 2 provides a summary of several key studies where topic modeling was utilized to analyze textual data and generate insights within the IS field and related disciplines. These examples, published in top-tier journals and conferences, showcase the versatility of topic modeling in addressing a wide range of research questions, from understanding industry dynamics to analyzing user-generated content on various online platforms.

Author(s) & Year	Research Context	Data Source	Application of Topic Modeling
Y. Yang & Subramanyam (2023)	Online Knowledge Communities & Consumer Reviews	Four textual datasets, including an online knowledge community and online consumer reviews.	Proposed and validated StableLDA, a new method to address the instability of standard LDA models, demonstrating its ability to produce more consistent estimations for econometric analysis.
Zhang et al. (2024)	Online Knowledge Communities (OKCs)	Datasets from Stack Exchange and Quora.	Developed TM-OKC, a novel topic model that explicitly accounts for the structural and temporal dependencies between questions and answers in OKCs, showing its superiority over benchmark models for user profiling.
Shi et al. (2016)	Industry Intelligence & Competitive Analysis	Unstructured text from firms' business descriptions (e.g., 10-K filings).	Used topic modeling to construct a novel, fine-grained measure of business proximity between firms, validating its effectiveness in an analysis of mergers and acquisitions.
Namvar et al. (2022)	Electronic Word of Mouth (eWoM) & Public Health Tech	User reviews of Contact Tracing Mobile Applications (CTMAs) during the COVID-19 pandemic.	Developed a method for iterative seed word generation to guide interactive topic models, showing the approach could explain user satisfaction more effectively than purely unsupervised methods.
Vakulenko et al. (2014)	Mobile Application Ecosystems	Nearly 600,000 English-language app descriptions from the iTunes App Store.	Applied LDA to bootstrap the app categorization process, demonstrating that the automatically generated topics could complement, specify, and extend the existing manual categories.
Hannigan et al. (2019)	Management & Political Science	Media statements about British parliament members.	Used topic modeling to generate new conceptual linkages, specifically analyzing how the appearance of certain topics in media statements impacted the departures of politicians.

Table 2. Applications of Topic Modeling in IS and Management Research

While standard LDA is a powerful tool, it has a major problem with getting different results each time it is run, meaning that using the same model on the same data can produce results that are not the same and cannot be repeated (Y. Yang & Subramanyam, 2023). This inconsistency can weaken the trustworthiness of research findings, especially when the topic model's outputs are used for later statistical analysis. To solve this important issue, this study adopts StableLDA, which reduces this inconsistency by using groups of related words from word embeddings, to guide the model to produce more stable results (Y. Yang & Subramanyam, 2023). This method has been shown to greatly improve model stability while keeping the topics just as good or even making them better, leading to more dependable results in later analyses. By choosing StableLDA, this study ensures that the factors found in worker reviews are strong and reliable, and that the following tests are based on a foundation that can be trusted and repeated.

Beyond classical LDA-family methods, transformer-based topic models such as BERTopic (Grootendorst, 2022) have emerged as prominent alternatives. BERTopic generates topics by embedding documents with pre-trained language models, reducing dimensionality through Uniform Manifold Approximation and Projection (UMAP), clustering the embeddings with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), and deriving topic representations via class-based Term Frequency–Inverse Document Frequency (TF-IDF). Because BERTopic preserves contextual semantics through dense embeddings rather than bag-of-words co-occurrence, it can yield topics that are highly semantically coherent and has shown advantages in short-text and exploratory clustering settings (Egger & Yu, 2022). Other neural topic models, such as Top2Vec (Angelov, 2020) and the Embedded Topic Model (Dieng et al., 2020), similarly rely on embedding-based text representations. When dense semantic clustering or exploratory interpretation are the primary objectives, BERTopic and related neural topic models may be appropriate choices.

The choice of topic model in this study, however, is guided by alignment with the broader research design rather than by absolute model performance. Two design requirements are particularly important. First, the

downstream econometric analysis treats topic probabilities as document-level features, requiring a mixed-membership representation in which each review is expressed as a probability distribution over topics. This representation is intrinsic to LDA-family models but is not the default output of BERTopic, which assigns each document to a single cluster through HDBSCAN. Second, because instability in topic probabilities can propagate into the coefficients of subsequent regression models, the topic model must yield reproducible variables across runs. This is precisely the stability gap that StableLDA was developed to address (Y. Yang & Subramanyam, 2023). For the hybrid theory-building and theory-testing design adopted in this thesis, StableLDA therefore aligns most closely with the analytical requirements of producing stable, mixed-membership topic representations suitable for both inductive interpretation and downstream econometric estimation.

2.5 Deep Learning Model

In addition to the LDA method, this study also uses deep learning to develop a predictive model. This bridges the gap between the inductive theme discovery from topic modeling (Section 2.4) and the development of a high-performance predictive model. While topic models such as Latent Dirichlet Allocation (LDA) are effective for identifying emergent themes from a corpus (Berente et al., 2019), they are methodologically limited. These models operate on a "bag-of-words" principle, which captures the frequency of co-occurring words (Blei et al., 2002) but largely fails to incorporate semantic context, grammar, or nuanced meaning. To overcome these limitations and build a robust predictive model, this research employs deep learning, specifically pre-trained transformer models.

The methodological approach of this thesis is situated within the "text-as-data" paradigm, a core component of modern computational social science (CSS). This paradigm repurposes the textual artifacts of social life, such as reviews, legislative debates, or social media posts, as large-scale, structured data amenable to quantitative analysis.

The first major wave of computational text analysis in social science was driven by probabilistic topic models, most notably LDA (Blei et al., 2002). As discussed in the previous section, the primary strength of LDA lies in its utility as an inductive tool. By clustering co-occurring words into "topics", it allows researchers to discover and interpret the latent thematic structures within a corpus, thereby supporting the data-driven development of new theories.

The second wave arrived with the advent of deep learning, specifically pre-trained language models (LLMs) based on the transformer architecture, such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019). Transformers revolutionized Natural Language Processing (NLP) by introducing a mechanism for "contextual" embeddings; the meaning of a word is dynamically defined by the words surrounding it.

This methodological shift presented a significant challenge for its adoption in computational social science: the interpretability-performance trade-off. While topic models like LDA are highly interpretable and thus useful for explaining and testing social phenomena, their predictive performance is often weak. Conversely, high-performance transformer models are notoriously "black boxes," making it difficult to trace how they arrive at a prediction, which can be a barrier to their use in theory-testing (Grossmann et al., 2023).

More recently, the rapid emergence of even larger, generative LLMs (e.g., the GPT-family) has introduced a new methodological consideration for CSS researchers: the trade-off between model maturity and the state-of-the-art. When adopting any new computational method, social science as a discipline must move cautiously, prioritizing the validity, reliability, and methodological soundness of a tool before its widespread adoption (Lazer et al., 2020). While newer generative models are powerful, research indicates they are not a one-size-fits-all solution. For the common CSS tasks of classification and regression where labeled data is available (as in this thesis), domain-adapted BERT-family models remain superior, more mature, and more effective. Furthermore, CSS research must balance state-of-the-art performance with practical constraints of computational cost, data availability, and model maturity. Large generative models are computationally expensive and can be less straightforward to adapt for specific downstream tasks like

classification. This methodological choice mirrors the continued popularity of topic modeling; while "older," the BERT/RoBERTa family of models is mature, well-understood, and highly effective for academic research, particularly in resource-constrained environments.

Therefore, this study selected the DistilRoBERTa model (Sanh et al., 2020). As a distilled version of the RoBERTa model, it is specifically trained to be smaller, faster, cheaper, and lighter. It achieves this efficiency while retaining the vast majority of the original model's high performance, representing an optimal balance of predictive power and research-cycle feasibility.

There are two primary methodologies for applying transformers: fine-tuning the entire model or using it as a static feature extractor to generate embeddings. This study employs the latter "embeddings-as-features" approach. This study utilizes the pre-trained DistilRoBERTa -base model to convert the raw Pros and Cons text from each worker review into high-dimensional numerical vectors, or "embeddings". Specifically, the model processes the text and extracts the final hidden-state vector of the special <s> (classification) token. This <s> embedding is designed to serve as an aggregate semantic representation of the entire text sequence. The resulting 768-dimension vector provides a rich, semantically aware set of features for the downstream regression model.

Chapter 3. Methodology

To identify factors influencing worker perception of pay fairness on microtask crowdsourcing platforms, MTurk workers are selected as a representative sample of microworkers for the empirical study. Since the MTurk platform lacks a built-in review function or reputation system, self-organized online communities and third-party platforms have emerged, allowing workers to review requesters and share their lived experiences of working on MTurk (Irani & Silberman, 2013, 2016). Leveraging publicly available worker-generated reviews, this study identifies factors underlying workers' experiences by applying topic modeling techniques to qualitative reviews regarding workers' positive experiences. This study further examines the impact of these factors on the worker perception of pay fairness through econometric modeling and deep learning models.

As online reviews are posted voluntarily by workers in an unobtrusive manner, they provide rich narratives of microwork experiences. Although such reviews may be polarized (Schoenmueller et al., 2020), this polarity allows for capturing extreme opinions (e.g., perceptions of extremely fair or unfair pay) that are often underrepresented in survey data. Consequently, online reviews can complement surveys by revealing issues respondents may not recall or may be reluctant to disclose, for instance due to social desirability bias. This triangulation helps address biases in self-reports and strengthens causal inference about fairness antecedents and outcomes. Behavioral log analyses and platform-level audits provide objective evidence on time allocations, rejections, and realized wages (Berg et al., 2018; Hara et al., 2018), while controlled experiments isolate how pay levels, task meaning, and feedback influence effort, speed, and quality (Chandler & Kapelner, 2013; Mason & Watts, 2010). These complementary approaches support the study's hybrid, data-driven theory building and theory-testing design in the microwork setting. Figure 1 summarizes the hybrid methodological framework, linking the organizational justice foundation to the two analytical phases, inductive topic modeling for theory building and predictive modeling (OLS and DNN) for theory testing, that together operationalize perceived pay fairness.

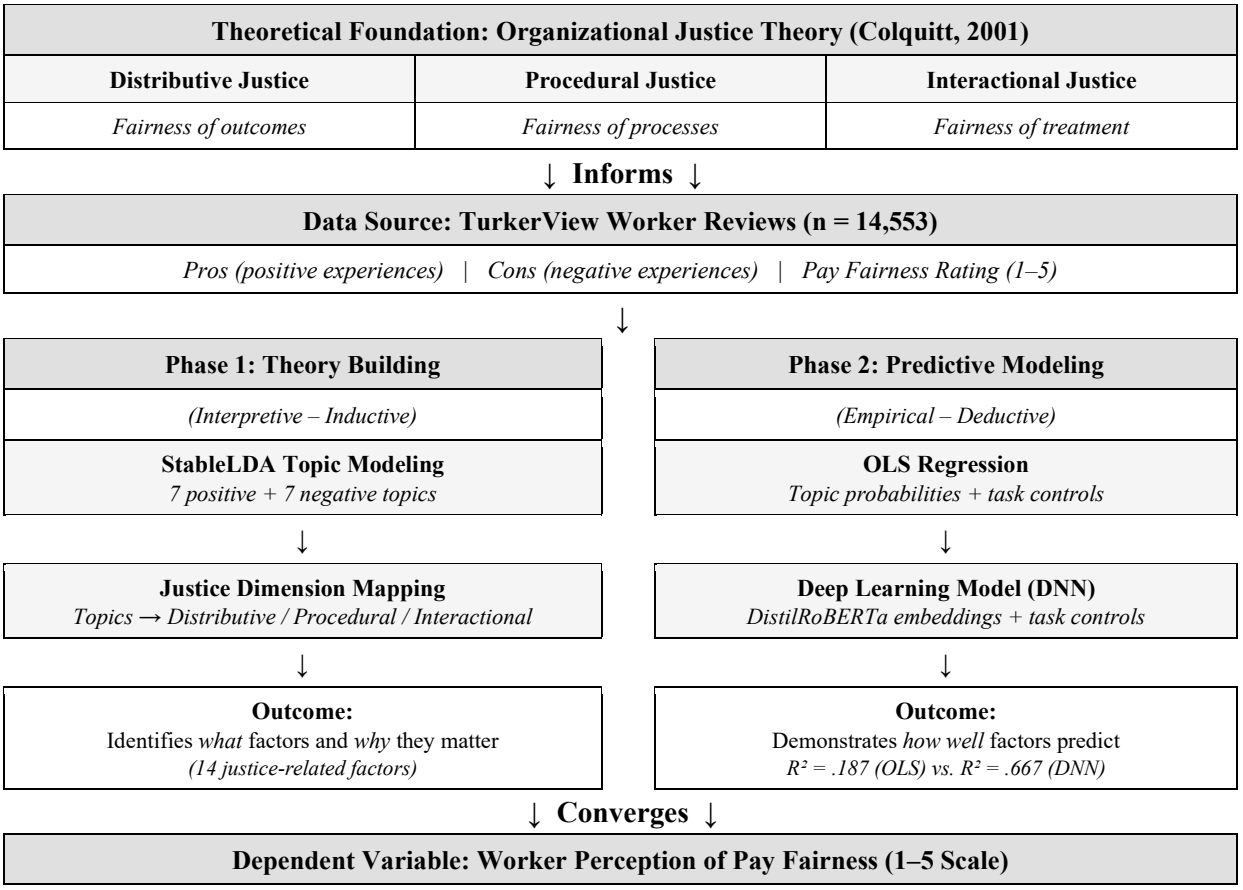

Figure 1. Hybrid Methodological Framework

3.1 Data Collection

This study collects data from TurkerView, a community-driven reputation platform designed to bridge the information gap between MTurk workers and requesters. Unlike general forums, TurkerView provides a structured review interface where MTurk workers log detailed qualitative and quantitative feedback metrics on completed HITs. Each review entry includes Wage Aggregates (calculated hourly rates based on completion time and reward), Reward Sentiment (a 1–5 Likert-scale rating ranging from "Underpaid" to "Generous"), and Communication Scores (rating requester responsiveness, not always available). Additionally, the platform aggregates individual worker statistics on profile pages, tracking their total reviews, approval ratings, and accepted work history.

Step 1: Data Source

TurkerView: Community-driven MTurk reputation platform
Review fields: Pros (text), Cons (text), Reward Sentiment (1–5 Likert scale)
Metadata: Completion Time, Task Payment, Approval Status, Worker Pace

↓

Step 2: Web Scraping

Custom Python scraper for TurkerView
169,415 reviews collected
September 2017 – February 2024 | 3,601 unique workers

↓

Step 3: Text Preprocessing (Gensim, Python)

(i) Lowercasing and tokenization
(ii) Removal of punctuation, URLs, and domain-specific stopwords
(iii) Lemmatization and stemming
(iv) Retention of currency and time expressions

↓

Step 4: Quality Filtering

Reviews containing fewer than five words excluded

↓

Final Analytical Corpus

n = 14,553 reviews
→ *Phase 1 (Topic Modeling) and Phase 2 (Predictive Modeling)*

Figure 2. Data Collection Process

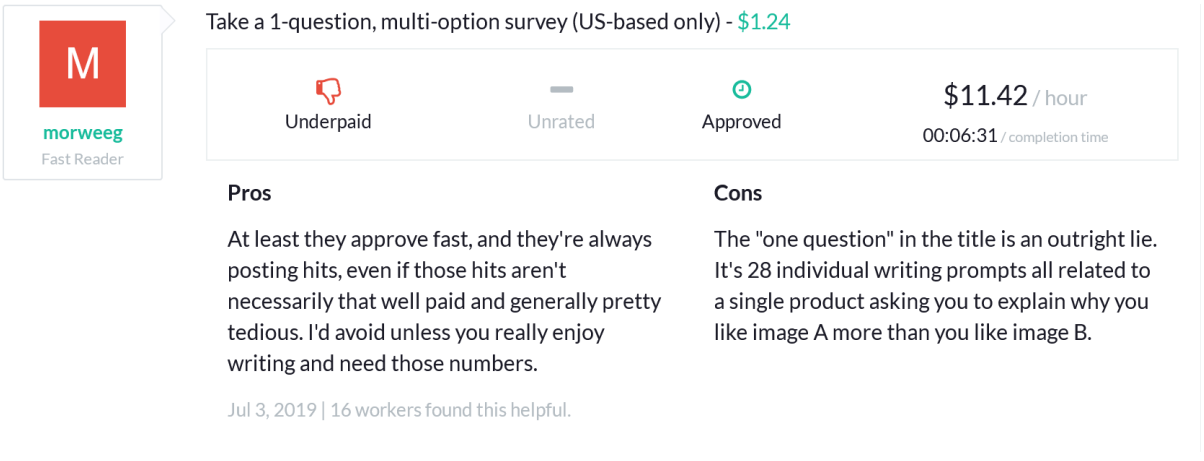

Figure 3. Sample TurkerView Worker Review

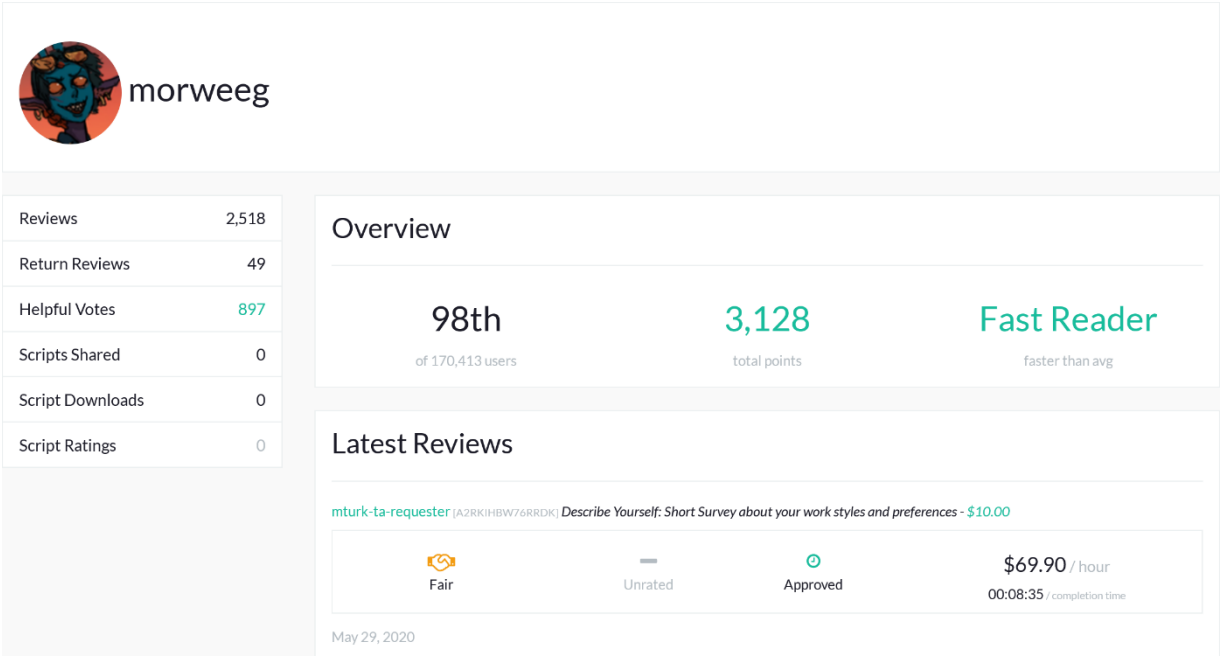

Figure 4. Sample TurkerView Worker Profile

Figure 2 illustrates the four-step data collection pipeline. Developing a web scraper for TurkerView, this study systematically collects 169,415 reviews posted between September 2017 and February 2024 by 3,601 unique MTurk workers. Figure 3 and Figure 4 show representative review and profile pages from which these data were extracted. The raw data is preprocessed using the Gensim library in Python, through lowercasing and tokenization, followed by the removal of punctuation, URLs, and high-frequency domain-specific stopwords. The text was then lemmatized and stemmed, and currency and time expressions were retained to capture compensation and effort. To reduce noise from uninformative reviews, those containing fewer than five words are excluded, resulting in a final corpus of 14,553 reviews. Overall, these preprocessing steps aim to enhance data quality without introducing bias into the results. Figure 5 reports the distribution of the Reward Sentiment ratings in the final corpus, showing the empirical spread of perceived pay fairness across the 14,553 analyzed reviews.

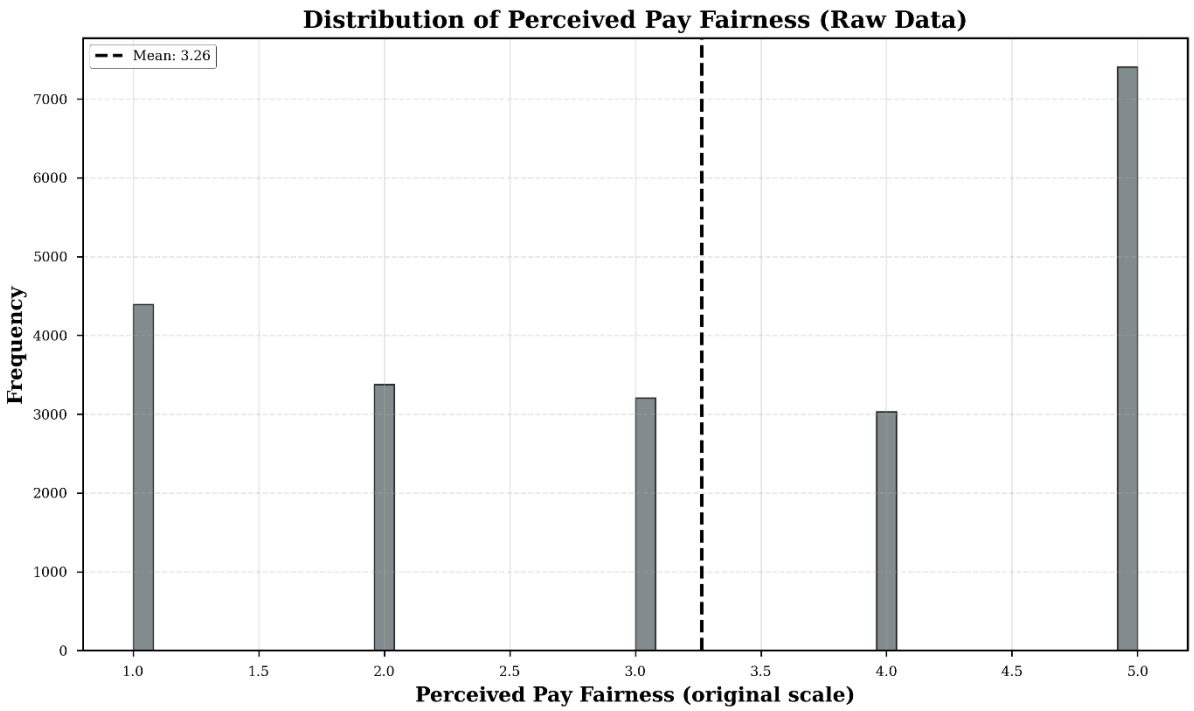

Figure 5. Distribution of Perceived Pay Fairness

3.2 Topic Modeling

Focusing on workers' fairness perceptions, this study analyzes the preprocessed textual data of their positive work experiences using topic modeling, an unsupervised machine learning technique for identifying latent themes (or topics) from a collection of text data (Shi et al., 2016). Latent Dirichlet Allocation (LDA), a widely used probabilistic model for topic modeling (Blei et al., 2002), was employed to extract latent topics within worker reviews. To determine the optimal number of topics for the LDA model, this study employs a comprehensive evaluation approach using five different metrics: Arun2010, Cao2009, Perplexity, Coherence_uci, and Coherence_cv. Each metric provides a unique perspective on the quality and interpretability of the topic model. These metrics are plotted and examined across different potential k values to derive the pattern of these five metrics. The number of topics that best balances the trade-off between within-topic coherence and between-topic exclusivity across models with varying topic numbers is selected from these patterns, with the resulting solution reported in Chapter 4. This study utilizes the

StableLDA method (Y. Yang & Subramanyam, 2023) for its consistency and robustness. It maintains consistent interpretations across multiple model-runs and produces more reproducible results for downstream econometric analysis. Based on the outputs of StableLDA (e.g., each topic's top terms and representative quotes), researchers identify meaningful topics, interpret the key theme represented by each topic, and assign informative labels to these topics accordingly.

The Arun2010 metric utilizes Jensen-Shannon (JS) divergence between the distribution of topics across documents and the distribution of words within topics to measure the similarity between topics, with lower values indicating better topic separation and a more distinct topic structure (Arun et al., 2010).

$$Arun2010(k) = \sum_{i=1}^{k} JS\big(p(\text{words}|t_i) \parallel p(\text{docs})\big)$$

Where:

$p(words|t_i)$ denotes the word distribution for topic t_i,

$p(docs)$ represents the empirical distribution of documents.

Similarly, the Cao2009 metric evaluates topic density and overlap among topics by measuring the cosine similarity between topic-word distributions, where lower scores suggest more distinctive and less overlapping topics (Cao et al., 2009).

$$Cao2009(k) = \frac{1}{k(k-1)} \sum_{i=1}^{k} \sum_{j \neq i} \cos\big(\beta_i, \beta_j\big)$$

Where β_i and β_j are the word distribution vectors for topic i and j.

Perplexity was also calculated to assess the model's predictive ability on unseen data, with lower perplexity values indicating better generalization and predictive capability (Blei et al., 2002).

$$Perplexity(D_{test}) = \exp\left(-\frac{\sum_{d=1}^{M} \log p\,(\boldsymbol{w_d})}{\sum_{d=1}^{M} N_d}\right)$$

Where:

M is the number of test documents,

N_d is the number of words in document d,

$p(\boldsymbol{w_d})$ is the probability of the words in document d under the model.

For measuring semantic interpretability, two coherence metrics are incorporated: $\text{Coherence}_{\text{UCI}}$, which uses sliding windows and pointwise mutual information (PMI) to evaluate top word co-occurrence within topics, and $\text{Coherence}_{\text{CV}}$, which combines pointwise mutual information (NPMI) based similarity with cosine similarity to assess the overall semantic coherence (Röder et al., 2015). Higher coherence indicates stronger semantic relatedness among top words, enhancing interpretability.

$$UCI(k) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \text{PMI}(w_i, w_j)$$

Where:

N is the number of top words per topic,

$\text{PMI}(w_i, w_j) = \log \frac{\text{P}(wi,wj)+\epsilon}{P(w_i)P(w_j)}$,

$P(w_i, w_j)$ is the joint probability of words w_i and w_j,

$P(w_i)$ and $P(w_j)$ are individual probabilities.

$$CV(k) = \frac{1}{N} \sum_{i=1}^{N} \cos\left(\boldsymbol{v}_{\boldsymbol{w_i}}, \sum_{j=1}^{N} \boldsymbol{v}_{\boldsymbol{w_j}} \right)$$

Where:

$\boldsymbol{v}_{\boldsymbol{w_i}}$ is the context vector for word w_i,

$\boldsymbol{v}_{\boldsymbol{w_j}}$ is similarly defined for other top words,

$\text{NPMI}(w_i, w_j) = \frac{\text{PMI}(wi,wj)}{-\log P(w_i,w_j)}$.

3.3 Predictive Modeling

To quantitatively test the factors influencing worker perception of pay fairness, this study implements and compares two distinct predictive models. The first model serves as a baseline, using the topic probabilities derived from Phase 1. The second employs a state-of-the-art deep learning architecture that models the raw text directly. Both models are trained to predict the same dependent variable: the worker's numerical rating of pay fairness on a 1-5 scale, referred to in the dataset as RewardSentiment.

3.3.1 Econometric Model

To further investigate the influence of work experience on worker perception of pay fairness, an econometric model is developed to predict workers' ratings of pay fairness using the factors derived from the topic modeling. For each textual review of completing a MTurk task, TurkerView also provides information on key task features and the characteristics of the worker who submitted the review, which are incorporated into the predictive model (see Equation 1).

$$\begin{aligned}\mathbf{PayFairness}_i = \boldsymbol{\beta}_0 + \sum_{k=1}^{N} \boldsymbol{\beta}_k \mathbf{Topic}_{i_k} + \boldsymbol{\beta}_{N+1}\mathbf{CompletionTime}_i \\ + \boldsymbol{\beta}_{N+2}\mathbf{TaskPayment}_i + \boldsymbol{\beta}_{N+3}\mathbf{ApprovalStatus}_i \\ + \boldsymbol{\beta}_{N+4}\mathbf{WorkPace}_i + \boldsymbol{\varepsilon}_i\end{aligned} \quad \text{Equation (1)}$$

The dependent variable, perceived pay fairness (PayFairness), is measured using a numerical rating on a 1 to 5 scale provided by reviewing workers on TurkerView, reflecting their perception of pay fairness for a HIT they completed on MTurk. This measure is recorded at the review level, representing a unique worker-task pair that captures an individual worker's evaluation of a specific task instance. The independent

variables include the factors listed in Table 3. The worker characteristics are included to account for potential individual differences in perceptions of pay fairness (Kim et al., 2019).

Variable	Description
Topic_{i_k}	Variables representing the probability that review i is associated with topic k (k=1 to N, optimal number of topics). These topic probabilities quantify the relative prominence of each justice-related factor within worker reviews.
CompletionTime_i	The actual time in seconds taken to complete the task as reported in review i.
TaskPayment_i	The effective hourly compensation (USD) calculated as task payment divided by completion time in review i.
ApprovalStatus_i	Binary variable indicating whether the task was approved (1) or rejected (0) in review i.
WorkPace_i	A numerical rating (on a scale of 1 to 7) representing the relative task completion speed of the worker who posted review i compared to that of other workers.

Table 3. Variables Used for Modeling Pay Fairness Perception

To assess the incremental explanatory power of text-derived topic factors over and above task-level controls, four nested OLS specifications are estimated in a stepwise fashion:

Model 1 (Controls Only) serves as the baseline, regressing perceived pay fairness on the four control variables only (CompletionTime, TaskPayment, ApprovalStatus, and WorkPace). This specification captures the variance attributable to observable, quantitative task and worker characteristics.

Model 2 (Controls + Pro Topics) augments the baseline with the seven pro-factor topic probabilities derived from the StableLDA analysis of positive reviews. This isolates the additional explanatory power contributed by the thematic content of workers' positive experiences.

Model 3 (Controls + Con Topics) augments the baseline with the seven con-factor topic probabilities instead. This isolates the additional explanatory power contributed by the thematic content of workers' negative experiences.

Model 4 (Full Specification) is the complete model, incorporating all four control variables alongside both pro-factor and con-factor topic probabilities. This full specification allows the joint contribution of all textual predictors to be assessed and provides the most comprehensive econometric account of the factors associated with perceived pay fairness.

This stepwise approach enables a transparent decomposition of explanatory power, revealing whether positive topics, negative topics, or their combination provide the greatest incremental improvement over quantitative controls alone. The comparison also serves as the econometric baseline against which the deep learning architecture is evaluated.

In addition to the pooled specifications, WorkPace-stratified OLS models are estimated to examine whether the associations between task characteristics and perceived fairness vary across worker experience levels. Workers are grouped into four WorkPace categories (High, Medium-High, Medium-Low, and Low) based on their WorkPace rating, and separate regressions are estimated within each stratum. This stratification allows the analysis to detect experience-related heterogeneity in how workers weight monetary returns, time investment, and qualitative task signals when forming fairness judgments.

3.3.2 Deep Learning Model

To move beyond thematic probabilities and capture the full semantic richness of the worker reviews, a custom Deep Neural Network (DNN) was developed. This model directly operationalizes the approach outlined in the literature review (Section 2.4), combining transformer-based text embeddings with structured numerical control variables.

The model's inputs are of two distinct types: textual inputs, comprising the raw strings from the Pros and Cons columns, and numerical inputs, consisting of the four aforementioned variables (CompletionTime, TaskPayment, ApprovalStatus, and WorkPace). The model architecture is composed of an embedding generation module followed by a feed-forward regression network. For embedding generation, the DistilRoBERTa -base transformer was used as a static feature extractor, processing the Pros and Cons text for each review independently. This process involved tokenizing the text (to a 512-token limit) and extracting the 768-dimension embedding of the <s> token from the model's last hidden state, yielding two 768-dimension vectors. These two text embeddings were then concatenated with the four scaled numerical variables, creating a single, fused input vector of 1,540 dimensions. This combined vector was then passed

into a 3-layer feed-forward network, structured as an input layer, three hidden layers of 256-128-64 neurons each with ReLU activation, and a final single-neuron output layer for the regression prediction. The model's training and evaluation were conducted after splitting the reviews into distinct training, validation, and test sets to ensure performance was measured on unseen data. The network was trained for 10 epochs using a batch size of 64, with all computations executed on an NVIDIA GeForce RTX 3090 GPU. The training process was optimized using the Adam optimizer with an initial learning rate of 1e-5, which was decayed by a factor of 0.1 every 5 epochs using a step scheduler. As this is a regression task, the model was trained to minimize the Mean Squared Error (MSE) Loss.

Chapter 4. Topic Modeling Results

This chapter presents the results of the topic modeling analysis. The chapter reports the optimal number of topics for workers' positive and negative reviews, and details the emergent topics and their alignment with organizational justice dimensions, respectively.

4.1 Topics of Positive Worker Experiences

The five model evaluation metrics are plotted across varying numbers of topics (i.e., k values) to assess model performance for workers' positive reviews (Pros). As shown in Figure 6, $k = 7$ emerged as the optimal number of topics, balancing model complexity and interpretability. Specifically, the Arun2010 and Cao2009 metrics showed declining trends that began to plateau around $k = 7$, indicating adequate topic separation with minimal redundancy. Perplexity continued to decrease but with diminishing returns beyond seven topics. Both coherence metrics (Coherence_UCI and Coherence_CV) exhibited favorable values at $k = 7$, suggesting strong semantic relatedness among top words within each topic. Considering the trade-off between model underfitting and overfitting, the seven-topic solution was selected for further analysis.

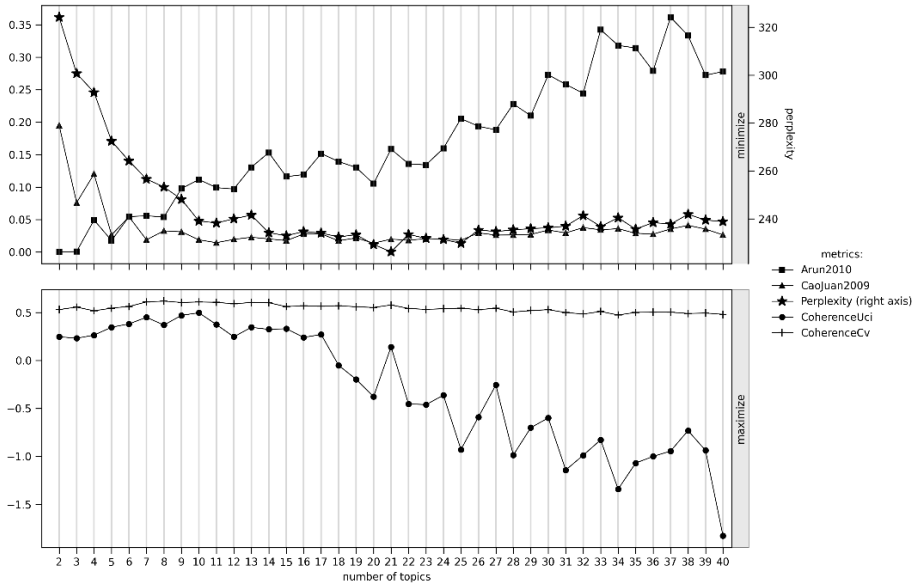

Figure 6. Model Evaluation Metrics for Topic Modeling (Pros)

The results of the seven-topic solution for pros are presented in Table 4, including the assigned topic labels based on thematic interpretation, the top 10 terms associated with each topic, and the prevalence of each topic based on its distribution across the dataset. These topics correspond to distinct aspects of work experience on microtask crowdsourcing platforms and are illustrated with representative quotes from the worker reviews. A full list of representative quotes for each positive topic is provided in Appendix A: Representative Quotes for Topics Emerging from Positive Reviews.

Topic Label	Top Terms	Topic Prevalence
1: Generous pay	pay, approv, good, hour, task, fair, rate, great, generous, instant	19.1%
2: Bonus opportunities	bonus, day, studi, base, receiv, payment, potenti, possibl, updat, cent	18.6%
3: Easy earnings	easi, quick, fast, write, work, pretti, simpl, game, requir, super	16.7%
4: Task Instructional Clarity	question, answer, read, demograph, scenario, instruct, type, relat, look, imag	12.4%
5: Time efficiency	time, minut, complet, timer, break, estim, long, take, probabl, state	13.0%
6: Requester responsiveness	request, reject, email, submit, respond, get, code, issu, give, say	12.9%
7: Well-structured task	survey, page, follow, prompt, bubbl, short, demograph, couple, video, listen	13.0%

Table 4. Results of Seven-Topic Solution for Pros

4.1.1 Topic 1: Generous Pay

This topic captures workers' satisfaction with monetary compensation that exceeds their expectations relative to effort expended. Workers value tasks where the payment is perceived as proportionate or generous compared to the expected rate. The top terms, including "fair", suggest that workers view generous payment rates as an important component of perceived pay fairness. A representative worker review within this topic reads: "Excellent overall pay for this study. This HIT accomplishes both having good pay and an excellent hourly rate, and that's nice!"

4.1.2 Topic 2: Bonus Opportunities

This topic demonstrates workers' gratitude for the opportunity to earn bonuses beyond the base rate as recognition for high-quality work or extra effort. The top terms such as "day" and "update" indicate that tasks offering bonuses involve follow-up components that facilitate bonus payments. The ability for workers to influence their earnings through their performance or active participation enhances their overall work experience. A representative review within this topic states: "In addition to the base payment, in each session you will have the opportunity to earn multiple additional bonus payments."

4.1.3 Topic 3: Easy Earnings

This topic reflects workers' positive perceptions of tasks that offer straightforward opportunities to earn income. In particular, the top terms highlight the game-like nature of certain tasks that are easy to complete and require minimal effort. A representative review within this topic notes: "super fast, designed for children so its pretty easy to understand and remember things."

4.1.4 Topic 4: Task Instructional Clarity

This topic features the informational clarity in workers' experiences, emphasizing the importance of well-articulated task instructions and clearly defined expectations. The top terms indicate these tasks typically involve reading scenarios, answering questions, and collecting demographic information. Task instructional clarity serves as an informative interface that facilitates effective interactions between workers and platforms, providing transparency regarding task requirements and expectations. A representative review within this topic describes: "Answer some societal type questions - read 5 different questions/scenarios answer each one using sliders - describe the questions."

4.1.5 Topic 5: Time Efficiency

This topic highlights time-efficient task execution as a key factor in how workers evaluate their experience with performing microtasks. The top terms indicate high time efficiency, particularly in tasks completed more quickly than initially estimated, creating a perception of "time surplus" that enhances their perceived value. A representative review within this topic points out: "Each block should take no more than 13 minutes (multiply by 6 blocks). Initial setup is no more than 5 minutes. Upload at the end took 4 minutes. Timer is sufficiently long."

4.1.6 Topic 6: Requester Responsiveness

This topic revolves around the interactions between workers and requesters, with a focus on the quality of communication between them. The top terms underscore that requester responsiveness, especially in

addressing issues or providing feedback on rejections, significantly influences workers' evaluations of their experiences. This demonstrates that while compensation plays a significant role, the quality of interpersonal treatment also has a substantial impact on workers' experiences in online workplace settings. A representative review within this topic depicts: "He was very communicative. I had emailed and he responded telling me what I had done wrong."

4.1.7 Topic 7: Well-structured Task

This topic focuses on the structural design and organization of microtasks. The top terms reflect workers' attention to specific task design features, such as question formats, page layouts, and navigation flows. It highlights the importance of task structures designed to support intuitive workflows, ensuring a smooth process that facilitates successful task completion and compensation. A representative review within this topic illustrates: "Hypothetical, followed by a few fill in the blanks. Short demographics at the end."

4.1.8 Positive Topic–Justice Alignment

Drawing upon organizational justice theory, the emergent topics are mapped to the well-established three-factor model of distributive, procedural, and interactional justice (Colquitt, 2001). In developing and validating this model, Colquitt (2001) generated items to represent each dimension, providing both conceptual definitions and operational measures. Each topic is then assigned to a dimension of organizational justice when its representative quotes fall within the scope of that dimension, using these definitions and measurement items as a coding scheme. This theory-driven mapping grounds the interpretations in established constructs and operational definitions, thereby enhancing the validity and rigor of the analysis.

Topic Label	Justice Dimension	Rationale for Alignment
1: Generous pay	Distributive Justice	Terms such as "pay," "rate," and "generous" explicitly refer to the financial outcomes and compensation received by the worker.
2: Bonus opportunities	Distributive Justice	Focuses on additional monetary rewards ("bonus," "payment," "cent"), directly impacting the perceived fairness of the economic outcome.

3: Easy earnings	Distributive Justice	Relates to the effort-to-reward ratio ("easy," "quick," "simple"), where low effort for the provided compensation is viewed as a favorable outcome.
4: Task Instructional Clarity	Interactional Justice	Centers on the quality of information provided ("instruction," "read," "question"), reflecting the clarity and explanation of the task requirements (Informational Justice).
5: Time efficiency	Distributive Justice	"Time" and "minute" represent the primary resource invested by the worker; efficiency dictates the "wage per hour," directly affecting the fairness of the final outcome.
6: Requester responsiveness	Interactional Justice	Involves communication and treatment by the requester ("respond," "email," "reject"), highlighting the interpersonal and informational aspects of the worker-requester relationship.
7: Well-structured task	Procedural Justice	Distinct from the outcome, this topic describes the mechanics and design of the task process ("page," "bubbles," "prompt"), which relate to the structural fairness of the workflow.

Table 5. Alignment of Seven-Topic Solution for Pros

Table 5 presents the alignment of these topics with the three dimensions of justice. Topics centered on compensation outcomes are mapped to distributive justice, which conceptually focuses on the fairness of outcome allocation. Specifically, generous pay (topic 1) directly reflects workers' overall evaluations of monetary compensation for microtasks, whereas bonus opportunity (topic 2) and easy earnings (topic 3) capture workers' perceptions of the effort-to-reward ratio. Time efficiency (topic 5) relates to the time invested relative to the compensation received. Procedural justice is manifested in well-structured task (topic 7), capturing workers' assessments of task design and workflows, which emphasize the processes and structural elements that shape how work is conducted.

Topics pertaining to workers' interactions with tasks and requesters align with the dimension of interactional justice. Specifically, task instructional clarity (topic 4) is indicative of the sub-dimension of informational justice given its focus on the clarity and adequacy of task-related information, whereas requester responsiveness (topic 6) reflects the quality of worker-requester interactions and corresponds to the sub-dimension of interpersonal justice. As topic modeling results include the probabilities of each review across the seven topics, this alignment not only allows for the quantification of each topic (or factor) but also provides an innovative approach to measuring justice perception.

4.2 Topics of Negative Worker Experiences

A parallel evaluation was conducted for workers' negative reviews (Cons). As shown in Figure 7, the five metrics exhibited similar convergence patterns, with $k = 7$ again emerging as the optimal solution. The consistency of this result across both positive and negative corpora provides additional confidence in the selected topic granularity. The seven-topic solution for negative reviews was therefore adopted for the subsequent analysis.

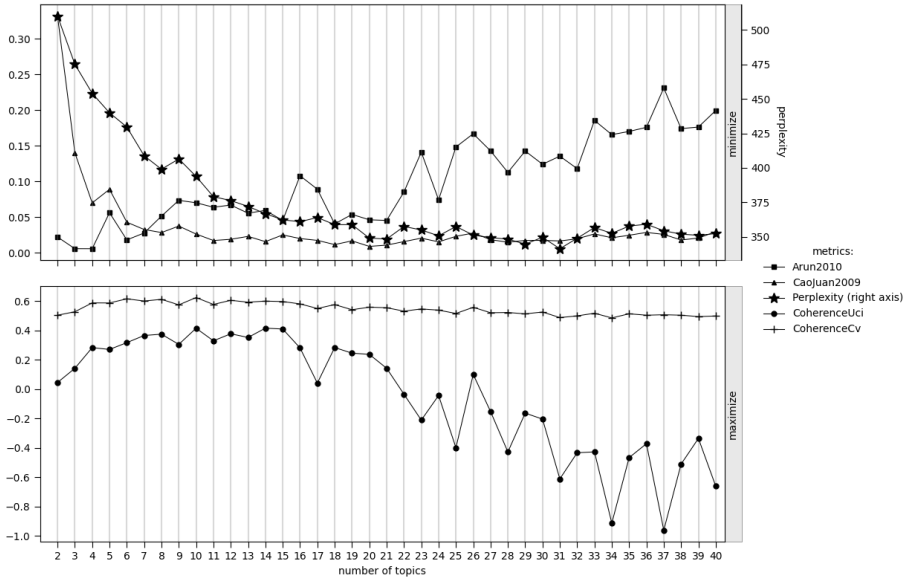

Figure 7. Model Evaluation Metrics for Topic Modeling (Cons)

Similarly, the results of the seven-topic solution for cons are presented in Table 6, also including the assigned topic labels based on thematic interpretation, the top 10 terms associated with each topic, and the prevalence of each topic based on its distribution across the dataset. These topics correspond to distinct aspects of work experience on microtask crowdsourcing platforms. Representative quotes are also selected from the worker reviews to show the corresponding aspects of these topics. A full list of representative quotes for each negative topic is provided in Appendix B: Representative Quotes for Topics Emerging from Negative Reviews.

Topic Label	Top Terms	Topic Prevalence
1: Task Interface Friction	underpaid like question way bad repetit feel bit make progress	21.3%
2: Excessive Task Length	write question bubbl read lot requir prompt word sentenc follow	19.4%
3: Technical Delays and Lag	time page minut long task take video pay timer short	16.2%
4: Task/Survey System Errors	survey code end complet submit give worker click mturk id	13.5%
5: Attention/Comprehension Check Failures	answer check attent think instruct sure reason miss correct ac	10.3%
6: Uncertainty in Earnings	bonus pay get actual base round tell cent low possibl	10.1%
7: Nonresponsive or Erroneous Rejections	reject request work studi respons say email approv receiv know	9.0%

Table 6. Results of Seven-Topic Solution for Cons

4.2.1 Topic 1: Task Interface Friction

This topic captures perceptions of poor effort-to-reward balance due to troublesome or confusing designs (for example, slider-heavy interfaces, misleading progress bars, or repetitive item phrasing). Workers describe tasks as unnecessarily effortful relative to the pay, which they interpret as unfair outcomes. A representative review states: “Steaming pile of garbage… $5 per hour… slider hell with an extremely inaccurate progress bar.”

4.2.2 Topic 2: Excessive Task Length

This topic reflects complaints about excessive length and content volume, especially large batteries of questions and multi-stage writing prompts that inflate completion time. The top terms emphasize writing, bubbles, and many questions. A representative review notes: “2 experiential writing prompts followed by tons of personality bubbles and product evaluations...”

4.2.3 Topic 3: Technical Delays and Lag

This topic highlights time lost to page loads, media buffering, and server lag that depresses effective hourly rate even when nominal pay appears acceptable. Workers emphasize delays that extend beyond task content

itself. A representative review explains: “Each audio clip takes 10-15 seconds to load… There were 80 clips in this HIT, so that's a lot of wasted loading time.”

4.2.4 Topic 4: Task/Survey System Errors

This topic centers on broken links, missing completion codes, and platform or survey integration errors that jeopardize payment despite completed work. These experiences are perceived as outcome unfairness because errors block compensation. A representative review warns: “the qualtrics page didn't generate a code… ‘Your Secret Code is CD Error: Dynamic Secret Parameters Missing (This message should not appear when your Qualtrics survey is viewed starting from a launched Mechanical Turk link)’”

4.2.5 Topic 5: Attention/Comprehension Check Failures

This topic captures frustration with attention checks and instruction quizzes that seem ambiguous, error-prone, or punitive. Workers feel unfairly failed even when they believe their answers were correct or consistent. A representative review states: “Failed the comprehension for the instructions check twice at 4/5… I'm pretty sure they have it set up so you auto fail regardless of your answers.”

4.2.6 Topic 6: Uncertainty in Earnings

This topic reflects tasks where earnings depend on random selection or other participants’ behavior, making outcomes feel uncontrollable or arbitrary. Workers object to bonus payments determined by chance or partner choices across multiple rounds. A representative review notes: “Waiting to get connected to a partner… 30 rounds… your bonus is based on one of these values… you could essentially do everything correct yourself and have a 50% chance of getting screwed if your partner assigns stupid values.”

4.2.7 Topic 7: Nonresponsive or Erroneous Rejections

This topic captures rejections perceived as mistaken, opaque, or unresponsive, including automated denials and accusations of duplicate participation. These experiences emphasize the interpersonal aspect of fairness in how requesters communicate and rectify errors. A representative review recounts: “EDIT: So I was

rejected for ‘work not found’… reached out to the requester… It’s been about 4 days now… I have yet to hear back… I won’t hold my breath.”

4.2.8 Negative Topic–Justice Alignment

Topic Label	Justice Dimension	Rationale for Alignment
1: Task Interface Friction	Distributive Justice	Terms like “underpaid,” “bad,” and “repetitive” highlight an imbalance where the effort required exceeds the compensation provided, a core distributive concern.
2: Excessive Task Length	Distributive Justice	The volume of work (“write,” “read,” “lot,” “sentences”) implies a high cost of effort that likely diminishes the effective value of the reward.
3: Technical Delays and Lag	Distributive Justice	Focuses on the duration (“time,” “long,” “timer”) required to complete a task; excessive time investment directly reduces the hourly wage rate (outcome).
4: Task/Survey System Errors	Distributive Justice	Technical failures (“code,” “submit,” “end”) often prevent workers from submitting work they have already completed, resulting in a total loss of the expected outcome (pay).
5: Attention/Comprehension Check Failures	Procedural Justice	Terms like “check,” “attention,” and “instruct” refer to the mechanisms used to validate work (e.g., attention checks), relating to the fairness of the evaluation process itself.
6: Uncertainty in Earnings	Distributive Justice	References to “bonus,” “low,” and “possible” suggest that the reward is contingent on chance or opaque variables rather than effort, affecting the certainty of the outcome.
7: Nonresponsive or Erroneous Rejections	Interactional Justice	Deals with the communication of negative feedback (“reject,” “email,” “say,” “respond”), focusing on how the requester treats the worker during disputes or denials.

Table 7. Alignment of Seven-Topic Solution for Cons

As summarized in Table 7, these topics for cons also map onto distributive, procedural, and interactional justice. Most cons reflect distributive justice concerns about unfair outcomes, particularly when effort is high and effective hourly pay is low due to Task Interface Friction (Topic 1), Excessive Task Length (Topic 2), Technical Delays and Lag (Topic 3), or Task/Survey System Errors (Topic 4). Procedural justice is captured by Attention/Comprehension Check Failures (Topic 5), where fairness of processes including attention checks, instruction quizzes, and validation rules comes into question. Interactional justice appears in Nonresponsive or Erroneous Rejections (Topic 7), where communication quality and responsiveness to errors shape perceived fairness. Uncertainty in Earnings (Topic 6) are mapped to distributive justice because

randomness and partner dependence affect final pay outcomes. The prevalence distribution indicates that negative worker experiences are dominated by distributive concerns, with procedural and interactional issues also playing critical roles in shaping perceived pay fairness on microtask platforms.

Chapter 5. Predictive Modeling Results

5.1 Econometric Models

Table 8 reports pooled normalized OLS estimates for four model specifications of increasing complexity. Model 1 serves as the baseline, regressing perceived pay fairness on the control variables only (CompletionTime, TaskPayment, ApprovalStatus, and WorkPace). Model 2 augments the baseline with pro-factor topic probabilities derived from the StableLDA analysis. Model 3 augments the baseline with con-factor topic probabilities instead. Model 4 is the full specification, incorporating both pro-factor and con-factor topic probabilities alongside all control variables. This stepwise approach allows the incremental contribution of each set of textual predictors to be assessed against the baseline, showing stable positive associations for effective payment and improvements in fit once topics are included. Selected topic coefficients are large and precisely estimated in the full model, consistent with this study's mapping from textual themes to organizational justice dimensions and the hypothesized role of qualitative signals in fairness judgements. Figure 8 plots the Model 4 coefficients with 95% confidence intervals, allowing direct visual comparison of the magnitude and precision of each topic's association with perceived pay fairness.

Variable	Model 1	Model 2	Model 3	Model 4
CompletionTime	-0.262***	-0.320***	-0.257***	-0.294***
	-0.045	-0.048	-0.047	-0.047
TaskPayment	0.325***	0.350***	0.330***	0.339***
	-0.083	-0.081	-0.087	-0.083
ApprovalStatus	-0.028*	-0.02	-0.034**	-0.026*
	-0.016	-0.016	-0.016	-0.016
WorkPace	0.226***	0.176***	0.220***	0.181***
	-0.018	-0.017	-0.017	-0.017
Pros T1		0.244***		0.236***

		-0.02		-0.021
Cons T1			0.037*	0.078***
			-0.021	-0.021
Pros T2		0.229***		0.233***
		-0.017		-0.017
Cons T2			0.067***	0.057***
			-0.022	-0.022
Pros T3		0.184***		0.186***
		-0.017		-0.017
Cons T3			0.126***	0.125***
			-0.021	-0.02
Pros T4		0.180***		0.181***
		-0.02		-0.02
Cons T4			0.067***	0.091***
			-0.02	-0.019
Pros T5		0.140***		0.131***
		-0.019		-0.02
Cons T5			0.136***	0.096***
			-0.022	-0.022
Pros T6		0.108***		0.109***
		-0.018		-0.019
Cons T6			0.085***	0.082***
			-0.021	-0.021

Observations	3425	3425	3425	3425
R-squared	0.111	0.175	0.127	0.187
Adj. R-squared	0.11	0.173	0.125	0.183
F Statistic	64.873	70.417	34.888	51.116

Table 8. Pooled OLS Regression Results (Model 1-4)

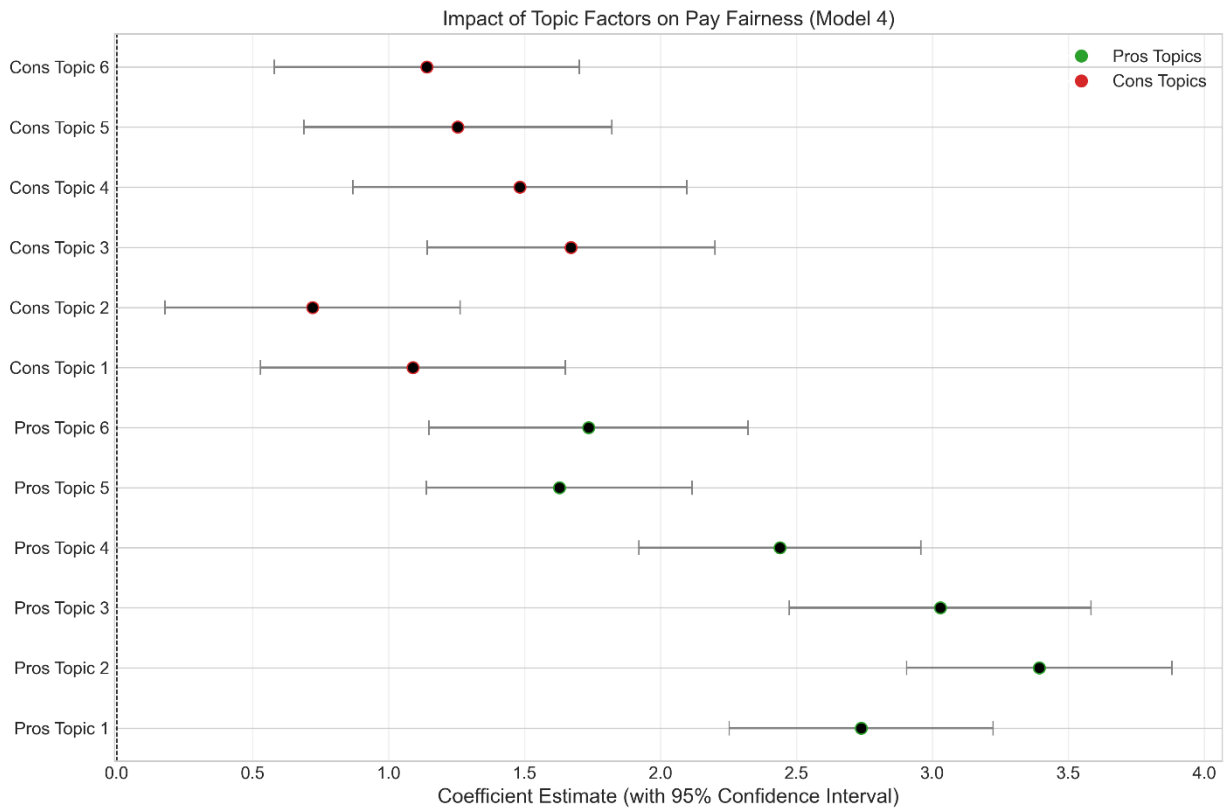

Figure 8. Coefficient plot (95% CIs) for Model 4

WorkPace-stratified OLS (Table 9) indicates that the marginal association of payment with fairness is largest in the Medium-Low and High pace groups, while the "within-stratum pace" term turns negative in the fastest group, suggesting distinct evaluation patterns among highly experienced workers. These differences reinforce the interpretation that experience shapes how workers integrate time, pay, and qualitative requester/task signals into perceived fairness.

Pace Category	CompletionTime	TaskPayment	ApprovalStatus	Pace	N	R^2
Low Pace	-0.1011 (0.0605)	0.195 (0.091)	-0.058 (0.030)	0.139 (0.038)	902	0.271
Medium-Low	-0.4718 (0.0566)	0.455 (0.059)	-0.023 (0.029)	-0.000 (0.000)	1136	0.153
Medium-High	-0.3232 (0.1058)	0.383 (0.083)	-0.080 (0.041)	-0.000 (0.000)	566	0.123
High Pace	-0.4280 (0.0747)	0.461 (0.071)	-0.024 (0.033)	-0.099 (0.036)	821	0.169

Table 9. Pace Stratified OLS Results

5.2 Deep Learning Models

To assess the most effective method for predicting a worker's perceived pay fairness, a comprehensive comparison was conducted between the OLS baseline model (regressing on topic probabilities) and four distinct Deep Neural Network (DNN) configurations. These DNNs utilize a standard multi-layer perceptron (256-128-64) but vary in their upstream sentence-to-embedding models: DistilBERT, DistilRoBERTa, RoBERTa, and MPNet. After training, their performance is evaluated using the same test set with R^2, which quantifies the proportion of variance explained (higher is better), as well as Mean Squared Error (MSE), which measures prediction error (lower is better). As illustrated in Table 10, DNN models dramatically outperform the OLS baseline on all predictive accuracy metrics.

Model	Model Description	R-squared (R^2)	MSE	Params (Language model)
DL DistilRoBERTa	Best balance of speed/accuracy.	0.667	0.807	~82M
DL RoBERTa	Larger architecture, slight overfitting.	0.650	0.848	~125M
DL DistilBERT	Older generation architecture.	0.605	0.957	~66M
DL MPNet	State-of-the-art sentence embedding model, specialized for semantic similarity.	0.568	1.050	~110M
OLS Model 4	OLS regression on StableLDA topic probabilities + numerical controls.	0.187	1.587	0

Table 10. Performance Comparison across Models

Figure 9, Figure 10, and Figure 11 visualize this gap from complementary angles: Figure 9 compares models by R^2, Figure 10 by residual distribution, and Figure 11 by the density of predicted versus observed fairness ratings. Together, they confirm that the DNN configurations and DistilRoBERTa in particular not only explain more variance but also produce tighter, better-calibrated predictions than the OLS baseline.

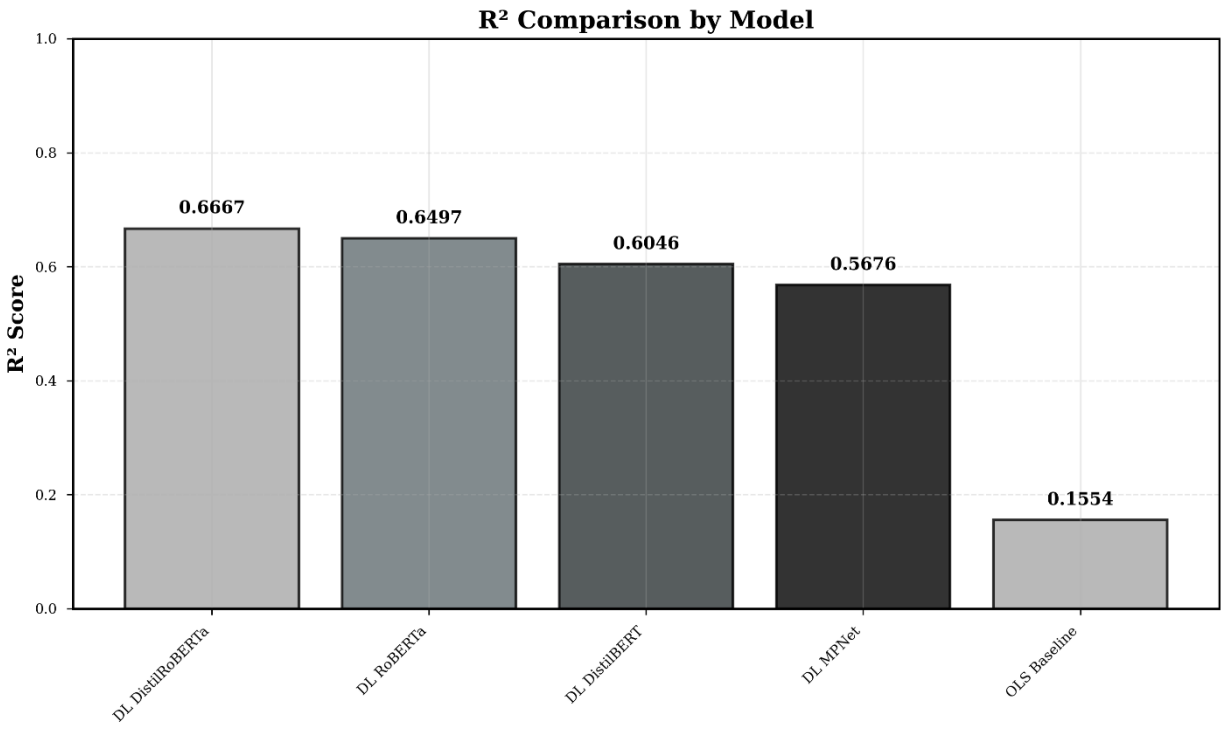

Figure 9. Model Comparison by R^2

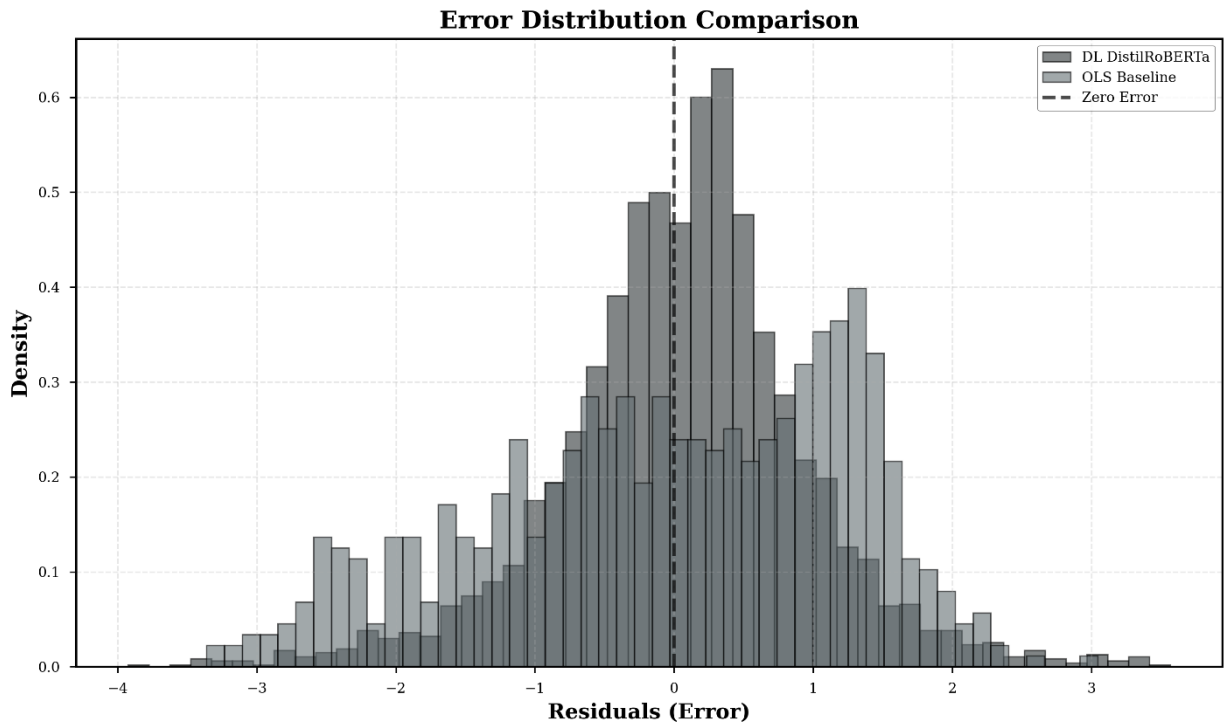

Figure 10. Model Comparison by Residuals

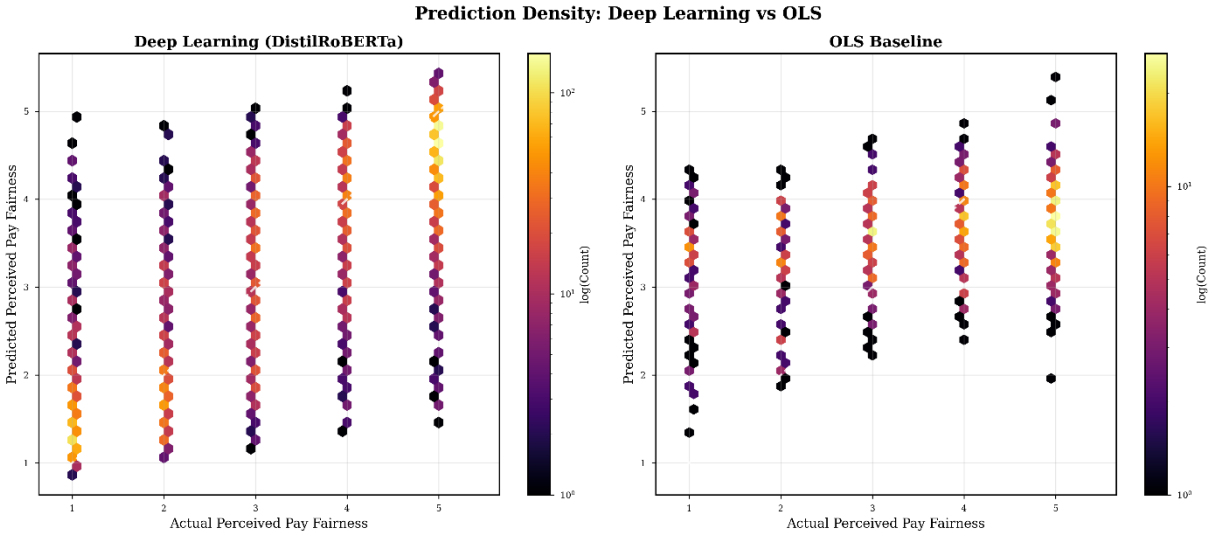

Figure 11. Model Comparison by Prediction Density

DistilRoBERTa emerges as the strongest model, achieving the highest R^2 (0.667) and the lowest MSE (0.807). It effectively balances parameter efficiency (~82M parameters) with generalization capability. RoBERTa, despite being a larger and more powerful model (~125M parameters), slightly underperformed its distilled counterpart ($R^2 = 0.650$, MSE = 0.848), suggesting that the larger architecture may have begun to overfit the training data, capturing noise rather than generalizable signal for this particular task.

DistilBERT, an older-generation architecture, achieved moderate performance ($R^2 = 0.605$, MSE = 0.957). Interestingly, MPNet, a current state-of-the-art semantic embedding model widely used in retrieval-augmented generation (RAG) tasks, performed the weakest among the four DNN configurations ($R^2 = 0.568$, MSE = 1.046). Its embedding space appears less optimized for the shorter, domain-specific textual reviews characteristic of crowdsourcing platforms. Based on these results, DistilRoBERTa is selected as the primary deep learning model for this study, offering the highest explanatory power and the lowest error rates.

5.3 Comparison of Econometric and Deep Learning Approaches

This section compares the predictive performance of the best-performing econometric specification (OLS Model 4) with the best-performing deep learning model (DNN with DistilRoBERTa embeddings). Table 11 consolidates the key performance metrics across both modelling paradigms. For comparability, the OLS baseline row reports the performance of an OLS regression on topic probabilities and numerical controls evaluated on the same held-out test set used for the deep learning models.

Model	R^2	MSE	RMSE	MAE
OLS Model 4 (Full)	0.187	1.587	1.258	1.049
DL DistilRoBERTa	0.667	0.807	0.898	0.663
OLS Baseline (topics only)	0.155	1.666	1.291	1.080

Table 11. Cross-Paradigm Performance Comparison

The most significant finding relates to the models' comparative explanatory power, quantified by the R^2 value. Leveraging semantic embeddings, this model successfully explained 66.7% of the variance in workers' perceived pay fairness ratings. This stands in stark contrast to the OLS baseline, which, using

topic probabilities, explained only 18.7% of the variance. This increase in explanatory power strongly indicates that the information contained in the raw text, when processed by a language model, is substantially more predictive than the topic model's high-level thematic abstractions. Beyond explanatory power, the deep learning model's predictions were also significantly more accurate in terms of error. The model's Root Mean Squared Error (RMSE) of 0.898 was 29% lower than the OLS model's 1.258, and its Mean Absolute Error (MAE) of 0.663 was 37% lower than the OLS model's 1.049. In the context of the 1-5 fairness rating scale, an average prediction error of 0.66 (DNN) versus 1.08 (OLS) represents a substantial improvement in practical, real-world accuracy.

These results highlight the complementary strengths of each modelling approach. The OLS model, which relies on topic probabilities, has difficulties capturing the majority of the variance in fairness perceptions. This suggests that while OLS combined with topic modeling is an excellent method for qualitatively identifying and interpreting the factors that workers discuss, it serves as an information-poor approach for building a robust predictive model. The process of abstracting rich, nuanced reviews into a small set of topic probabilities results in a significant loss of predictive information. The DNN, in contrast, proves highly effective.

Together, these results position the DistilRoBERTa-based DNN (selected in Section 5.2) and the OLS Model 4 (the strongest econometric specification) as complementary rather than substitutable: the OLS model offers interpretable coefficients anchored in the justice framework, while the DNN delivers substantially stronger predictive accuracy by preserving semantic signal that topic probabilities abstract away.

Chapter 6. Discussion, Contributions, and Implications

The rapid expansion of microtask crowdsourcing has reconfigured the structural and psychological boundaries of the traditional workplace. Within this highly mediated and volatile paradigm, evaluating how human workers perceive workplace fairness requires a profound synthesis of organizational justice theories, data mining, and computational analytics. The findings from this research bridge the methodological gap between the qualitative, subjective nuances of user-generated text and the requirements of quantitative predictive modeling. By isolating and analyzing these dimensions, this study demonstrates that justice perceptions in microtask crowdsourcing platforms are sensitive not only to the objective characteristics of the tasks such as baseline financial compensation and task duration, but also to the profound worker-requester interactions and algorithmic transparency. Ultimately, the following discussion explores how digital laborers formulate complex ethical evaluations of their invisible employers, highlighting the necessity for a reinterpretation of organizational justice theory to accommodate the realities of the microtask crowdsourcing platforms.

This chapter is organized as follows. The discussion begins by interpreting the topic modelling results, examining how the latent themes that emerged from workers' narratives align with and extend existing organizational justice frameworks. It then turns to the predictive modelling results, evaluating the comparative performance of econometric and deep learning approaches and drawing out the methodological and substantive implications of the model comparisons. Following these interpretive discussions, the chapter presents a summary of key findings before articulating the theoretical contributions and practical implications of the research.

6.1 Discussion on Topic Modeling Results

This study has generated seven topics each for positive and negative worker experiences on MTurk. Together, they create a comprehensive framework for understanding the multidimensional nature of workers' experiences on microtask platforms. While distributive justice factors dominate workers'

discussions, the procedural and interactional dimensions still represent significant portions of workers' positive evaluations. This distribution suggests that while fair compensation forms the foundation of worker satisfaction, the processes through which work is structured, and the quality of interactions also play crucial roles in shaping worker perception of pay fairness on microtask crowdsourcing platforms.

The topic modeling results are consistent with prior survey- and interview-based studies on crowdwork fairness while also revealing nuanced insights into previously examined factors. Although prior research has identified factors influencing perceived fairness in pay, such as remuneration, thoroughly explained task criteria, and communication between workers and clients (Alpar & Osterbrink, 2020; Fieseler et al., 2019; Schulze et al., 2023; Wood et al., 2019), the findings further uncover pay-related characteristics (e.g., generous pay, bonus opportunities, easy earnings) and task-related characteristics (e.g., time efficiency, well-structured task). Furthermore, the topic of time efficiency underscores the role of time in affecting perceptions of pay fairness, and the prominence of the term "game" within the topic of easy earnings signals its hedonic nature.

Integrating the positive and negative topic solutions under a common justice framework clarifies how each facet of the microwork experience maps onto organizational justice theory. Table 12 consolidates this mapping by putting all 14 topics (seven positives and seven negatives) against the three justice dimensions, together with each topic's prevalence in the corpus. There are some emerging patterns: First, distributive justice dominates both valences, accounting for four of the seven positive topics (67.4% cumulative prevalence) and five of the seven negative topics (80.5% cumulative prevalence); this asymmetry indicates that although outcome fairness is the primary lens through which workers evaluate their experience, concerns about unfair outcomes are even more concentrated on the negative side, where effort–reward imbalances surface through excessive task length, technical delays, system errors, and earnings uncertainty. Second, procedural justice appears in a more narrowly defined form in each valence: Well-structured Task on the positive side and Attention/Comprehension Check Failures on the negative side. This suggests that workers experience procedural fairness primarily through the design and enforcement of task validation

mechanisms. Third, interactional justice is represented on both sides (Task Instructional Clarity and Requester Responsiveness among positive worker experiences; Nonresponsive or Erroneous Rejections among negative worker experiences), underscoring that the quality of requester–worker communication shapes fairness perceptions regardless of valence. Taken together, Table 12 demonstrates that the seven-topic solutions recovered from naturalistic review text align coherently with the three-dimensional organizational justice framework while also revealing the relative weight that workers themselves place on each dimension in their unprompted accounts.

Justice Dimension	Positive Topics (Pros)	Negative Topics (Cons)
Distributive Justice *Fairness of outcomes*	**T1:** Generous Pay *(19.1%)* **T2:** Bonus Opportunities *(18.6%)* **T3:** Easy Earnings *(16.7%)* **T5:** Time Efficiency *(13.0%)*	**T1:** Task Interface Friction *(21.3%)* **T2:** Excessive Task Length *(19.4%)* **T3:** Technical Delays and Lag *(16.2%)* **T4:** System Errors *(13.5%)* **T6:** Uncertainty in Earnings *(10.1%)*
Procedural Justice *Fairness of processes*	**T7:** Well-structured Task *(13.0%)*	**T5:** Attention/Comprehension Check Failures *(10.3%)*
Interactional Justice *Fairness of treatment*	**T4:** Task Instructional Clarity *(12.4%)* **T6:** Requester Responsiveness *(12.9%)*	**T7:** Nonresponsive or Erroneous Rejections *(9.0%)*
Total	**7 topics (4 DJ + 1 PJ + 2 IJ)**	**7 topics (5 DJ + 1 PJ + 1 IJ)**

Note. DJ = Distributive Justice; PJ = Procedural Justice; IJ = Interactional Justice. Percentages indicate topic prevalence (proportion of the corpus associated with each topic). Topic numbers (T1–T7) refer to the numbering within each valence's seven-topic solution and are independently assigned.

Distributive justice dominates both positive (4/7, 67.4% cumulative prevalence) and negative (5/7, 80.5%) experiences. Procedural and interactional justice topics, though fewer in number, capture critical process- and communication-related fairness concerns.

Table 12. Topic-Justice Dimension Alignment

6.2 Discussion on Predictive Modeling Results

The predictive modelling results reveal a clear hierarchy in the capacity of different analytical approaches to capture the complexity of pay fairness perceptions. The econometric models, while providing interpretable coefficient estimates, demonstrated modest explanatory power, confirming the well-documented limitations of linear approaches in modelling subjective evaluations shaped by non-linear, interactive, and context-dependent factors. By contrast, the deep learning models achieved substantially higher predictive performance, with the combined text-embedding and task-control architecture explaining approximately two-thirds of the variance in fairness ratings. This marked improvement underscores a critical methodological insight: the semantic richness preserved in dense vector representations of worker narratives contains information about fairness perceptions that cannot be recovered through dimensionality-reducing techniques such as topic modelling. The attention mechanisms within the transformer architecture appear capable of identifying subtle linguistic patterns that workers employ to communicate the intensity and nuance of their fairness evaluations. These patterns are precisely the type of signal that traditional feature-engineering approaches fail to capture.

The WorkPace-stratified analysis further enriches the interpretation of the predictive modelling results. The observation that model performance varies across worker experience levels suggests that the linguistic strategies used to express fairness perceptions are not uniform across the workforce. More experienced workers, who have developed platform-specific vocabulary and evaluative frameworks, may produce reviews that are more semantically consistent and thus more predictable. Less experienced workers, by contrast, may exhibit greater variability in how they articulate fairness concerns. These findings underscore the importance of accounting for workforce heterogeneity in both predictive modelling and platform governance.

Taken together, the topic modelling and predictive modelling discussions reveal complementary insights: the former identifies the substantive dimensions of fairness that matter to workers, while the latter demonstrates that these dimensions are best captured through methods that preserve the full semantic richness of workers' language. The following section distils these interpretive discussions into a concise summary of the study's key findings before turning to broader theoretical and practical implications.

6.3 Summary of Key Findings

The regression evidence shows that higher effective payment is consistently associated with higher perceived pay fairness, while longer completion time is associated with lower perceived fairness, with pooled OLS estimates indicating a stable positive TaskPayment coefficient around 0.34 in the topic-augmented model and negative but small CompletionTime effects across specifications. Topic-augmented models substantially improve explanatory power over controls-only specifications, and the strongest topic signals align with generous pay, bonus opportunities, and easy earnings, indicating that justice-relevant textual cues shape fairness perceptions along with numeric task fields and worker pace.

The topic-justice mapping proposed in the thesis helps interpret these associations: distributive justice is reflected in topics emphasizing generous pay, opportunities for bonuses, easy earnings, and time efficiency; procedural justice is captured by well-structured task design; and interactional justice is reflected in clear instructions and requester responsiveness. The positive topic coefficients in the pooled models are theory-consistent with justice dimensions that predict evaluations of compensation as appropriate to effort and process, thereby connecting qualitative experiences in TurkerView reviews to quantitative fairness ratings at the review level.

Stratified models by worker pace reveal meaningful heterogeneity that helps explain fairness perceptions within experience segments of the labor market. The TaskPayment coefficient ranges from roughly 0.195 in Low Pace to about 0.455–0.461 in High and Medium-Low Pace groups, indicating that experienced or faster workers attach greater marginal weight to pay when forming fairness judgments, while the within-

stratum “pace” term turns negative only in the fastest segment, suggesting distinct evaluation patterns at the high-experience margin. These differences show that experience moderates reference points and time valuation, making workers’ fairness assessments more sensitive to monetary returns when pace and workflow mastery increase, while slower cohorts rely more on qualitative signals and incremental improvements in structure and communication for fairness judgments.

Another key finding of this thesis is the vast difference in predictive power between modeling methods. A worker’s 1-5 fairness rating is weakly predicted by the thematic topics they discuss (18.7% of variance explained) but strongly predicted by the semantic content of their raw text review when combined with task-level data in a deep learning model (66.7% of variance explained).

This implies that for understanding and predicting fairness, the way in which a worker describes their experience, including the specific sentiment, phrasing, and semantic nuance captured by the DistilRoBERTa-based embeddings is a far more critical signal than the general category of their complaint or praise (e.g., “Generous Pay” or “Rejection” as abstract topics).

6.4 Theoretical Contributions

This study provides several theoretical contributions to research on online labor markets, particularly pay fairness, by extending organizational justice theory from traditional workplaces to microtask crowdsourcing platforms, with a focus on the alignment between workers’ positive experiences and distributive, procedural, and interactional justice.

6.4.1 Reinterpretation of Justice Theory in Online Labor Markets

The primary theoretical contribution of this study involves the profound contextual reinterpretation of the Colquitt three-factor organizational justice framework to suit the realities of the gig economy (Colquitt, 2001). Each of the three dimensions of organizational justice has its own new meaning in microtask crowdsourcing platforms, developed with a novel two-phase hybrid approach. In Phase 1 (Interpretation),

an interpretable model (StableLDA) is first used to inductively build theory, identifying the key factors of fairness and mapping them to the Organizational Justice framework (Colquitt, 2001). This answers the "what" and "why" of fairness perception. In Phase 2 (Prediction), a high-performance, non-linear model leverages the raw data (text embeddings) to build an accurate predictive model. This demonstrates the practical utility of the identified factors, moving from theory-building to application.

By combining these two phases, the study reinterprets organizational justice theory for the specific context of online labor markets. Traditional justice theory was developed primarily in stable, hierarchical employment relationships where employer–employee interactions are ongoing and governed by formal contracts and institutional norms (Colquitt, 2001; Greenberg, 1990). In microtask crowdsourcing, however, work relationships are transient, anonymous, and mediated by digital platforms, fundamentally altering how each justice dimension manifests. The distributive dimension, for instance, is not merely about whether pay matches effort in a single job but about whether piece-rate compensation across many brief tasks accumulates into a livable effective hourly wage. Similarly, procedural justice in this context extends beyond formal grievance mechanisms to encompass the transparency and consistency of platform-mediated processes such as task qualification requirements, approval workflows, and rejection policies. Interactional justice, traditionally linked to supervisor–subordinate communication, is reframed around the quality and clarity of task instructions, the responsiveness of requesters to worker inquiries, and the provision of constructive feedback on completed work.

This reinterpretation enriches justice theory by demonstrating that its core dimensions remain operative in non-traditional employment settings, while also revealing that the relative salience and manifestation of each dimension shifts in response to the structural characteristics of platform-mediated work. The findings thus contribute to ongoing theoretical efforts to adapt organizational behavior frameworks to the evolving landscape of the gig economy and digital labor markets.

6.4.2 New Approach to Justice Measurement

The methodological contribution of this research is the development and theoretical validation of a novel, text-mining-based instrument designed to measure organizational justice perceptions using unstructured, user-generated content. Traditionally, the well-established approach for assessing organizational justice has been structured psychometric scales, most notably the four-dimensional scale developed by Colquitt (2001), which assesses distributive, procedural, interpersonal, and informational justice through Likert-type survey items. While these instruments have demonstrated robust construct validity in stable, traditional organizational settings, such as the automobile parts manufacturing plants where Colquitt's original validation occurred, their application to the volatile and fragmented context of crowdwork is constrained by methodological limitations. This study provides an alternative approach to measuring justice by operationalizing the construct directly from user-generated text, moving beyond traditional survey-based methods. This data-driven approach, in turn, provides enriched interpretations of justice in the specific context of online crowdwork. For instance, the analysis identified context-specific factors, such as time efficiency and ease of earning, that are not prominent in traditional justice literature but are central to shaping worker perceptions of pay fairness in this environment.

In the context of crowdwork, the administration of traditional surveys faces the critical obstacle of survey fatigue and the economic opportunity costs imposed on workers. Crowdworkers, who are typically compensated on a piece-rate basis, often view survey participation as unpaid or low-paid overhead that detracts from their effective hourly wage (Hara et al., 2018; Toxtli et al., 2021). This economic pressure encourages "satisficing" behavior, where workers may provide uniform or low-effort responses to minimize time expenditure, thereby compromising the reliability and validity of the data collected. Furthermore, the very act of administering a survey regarding fairness can induce a priming effect, framing the worker's subsequent cognitive processing and potentially inducing a rationalistic evaluation that divorces their reported perception from the spontaneous emotional states experienced during the actual work process

(Allport, 1954). The ecological validity of survey data is thus constantly in tension with the artificiality of the measurement tool itself, which separates the measurement of the experience from the experience itself.

To transcend these limitations, the proposed instrument leverages the ubiquity of textual data generated organically by workers within their own reviews. By treating this "found data" as a primary source, this study applies NLP techniques, specifically topic modeling methodologies like Latent Dirichlet Allocation (LDA) to extract justice dimensions directly from worker narratives. This approach aligns with the growing recognition in organizational research that text mining can accelerate knowledge discovery by radically increasing the volume of analyzable data while maintaining high ecological validity (Kobayashi et al., 2018). Recent empirical comparisons have demonstrated that semantic measures derived from open-ended text can yield statistical properties, such as validity and reliability, that are competitive with, or even superior to, traditional numerical rating scales, particularly when assessing psychological constructs like affect and well-being (Kjell et al., 2019).

Moreover, the application of this instrument addresses the black box problem often associated with machine learning in social science by grounding computational analysis in established organizational justice theory. By mapping the extracted topics back to the Colquitt three-factor model, this research bridges the gap between the computational power of data-driven analytics and the theoretical rigor of organizational behavior. This hybrid approach not only validates the existence of traditional justice dimensions in unstructured text but also allows for the detection of novel, context-specific factors that may not be captured by pre-defined survey items, such as the critical role of instruction clarity in shaping fairness perceptions. Consequently, this contribution answers the urgent call for methodological diversity in HRM research, moving beyond the Likert scale to capture the richness, complexity, and emotional depth of the gig worker experience.

Beyond introducing a text-mining-based measurement approach, this study also demonstrates the complementary value of combining an interpretable econometric model with a high-performing deep learning model. The topic-model-based predictive model offers theoretically meaningful and interpretable

evidence by estimating the relationship between specific justice-related factors and workers' pay fairness ratings, thereby allowing direct inference about which dimensions of organizational justice are most salient. In contrast, the DistilRoBERTa-based deep learning model captures richer semantic nuances in worker-generated text that extend beyond the explanatory scope of topic-level abstractions, resulting in substantially stronger predictive performance. Taken together, the two approaches illustrate that interpretability and predictive strength need not be viewed as competing priorities, but rather as complementary lenses for understanding the complexity of pay fairness perceptions in platform work.

6.4.3 Context-Specific Factors Influencing Pay Fairness

Several context-specific factors unique to microtask crowdsourcing warrant further discussion. The topic of time efficiency, which emerged as a positive theme, reflects a form of temporal economics particular to piece-rate digital labor: workers evaluate pay fairness not solely against the absolute monetary amount but against the ratio of compensation to time invested. This finding extends distributive justice theory by demonstrating that in microtask environments, the referent comparison is not merely "what others earn" but "what I could earn per unit of time across available tasks." The resulting calculus creates a dynamic, opportunity-cost-driven fairness assessment that differs fundamentally from the wage comparisons documented in traditional employment settings. The emergence of the easy earnings topic, characterized by hedonic language and frequent references to "game"-like experiences, suggests that perceived fairness in microtask work has an affective dimension that existing justice frameworks underspecify. When workers describe tasks as enjoyable or effortless, they are implicitly shifting their fairness referent from a purely economic exchange to a hedonic evaluation in which intrinsic satisfaction offsets lower nominal pay. This parallels findings in behavioral economics on the "effort-reward imbalance" model but extends it to digital piecework contexts where task variety and novelty serve as non-monetary compensatory mechanisms. Furthermore, the bonus opportunities topic represents a distinct fairness dimension that operates independently from base pay: workers appear to interpret bonuses not merely as additional income but as

signals of requester appreciation and recognition of quality, thereby engaging interactional justice perceptions alongside distributive ones.

The negative topics further illustrate context-specific procedural failures unique to platform-mediated work. System errors and technical delays, which would be classified as incidental inconveniences in traditional workplaces, assume justice-relevant significance in crowdsourcing environments because they directly reduce workers' earning capacity without recourse. Unlike salaried employees who are compensated regardless of system downtime, piece-rate crowdworkers bear the full cost of platform malfunctions. This structural vulnerability transforms routine technical issues into perceived procedural injustices, offering a context-specific extension of procedural justice theory that accounts for the distinctive risk allocation inherent in platform labor arrangements.

6.5 Practical Implications

The findings provide important practical implications for three key stakeholders in the crowdwork ecosystem: platform designers, requesters, and policymakers.

6.5.1 Implications for Platform Designers

For platform designers, the results underscore the importance of integrating mechanisms that promote fairness beyond pay, such as ensuring transparent processes, facilitating timely payment, and providing fair communication channels. The finding that procedural justice topics (such as well-structured tasks) significantly predict fairness perceptions suggests that investment in user interface design, standardized task templates, and algorithmic transparency can yield measurable improvements in workers' perception of fairness. The prominence of system errors and technical delays among negative topics further indicates that platform reliability is not merely a technical concern but a justice concern: workers perceive platform malfunctions as procedural failures that undermine the fairness of their work arrangements. Designers should therefore prioritize robust error-handling mechanisms, provide real-time status updates for task submissions, and implement transparent appeals processes for rejected work.

6.5.2 Implications for Requesters

For requesters, the analysis reveals that interactional and procedural justice are critical; requesters can improve fairness perceptions by communicating clearly, providing constructive feedback, and respecting worker effort and time. The results demonstrate that the quality of task instructions and the responsiveness to worker inquiries are not peripheral courtesies but substantive determinants of perceived pay fairness. The emergence of requester responsiveness and task instructional clarity as distinct positive topics indicates that workers evaluate the adequacy of their compensation partly through the lens of how they are treated during the work process. Requesters who invest in thorough pilot testing of task instructions, provide constructive feedback on rejected submissions, and maintain responsive communication channels can improve workers' fairness perceptions without necessarily increasing nominal pay rates. The bonus opportunity topic further suggests that even modest supplementary payments, when framed as recognition of effort or quality, can substantially enhance perceived distributive justice.

6.5.3 Implications for Policymakers

For policymakers, this study suggests that fairness standards must evolve beyond simple minimum pay rates to also promote process transparency and worker voice, addressing the procedural and interactional deficits common in platform work. The deep learning model's demonstration that semantic content predicts a substantial proportion of the variance in fairness ratings provides a scalable, automated instrument for monitoring worker welfare at the platform level. Regulatory agencies could leverage similar text-mining approaches to audit worker-generated reviews across platforms, identifying systemic fairness deficits without relying on costly and potentially biased survey instruments. Moreover, the finding that pace-stratified models reveal meaningful heterogeneity in fairness determinants across worker experience levels cautions against one-size-fits-all regulatory approaches: minimum wage mandates, while important, may be insufficient to address the procedural and interactional justice deficits that disproportionately affect less

experienced workers who lack the strategies and community knowledge to navigate platform-specific challenges effectively.

Chapter 7. Conclusions, Limitations and Future Work

7.1 Conclusions

This thesis set out to answer the primary research question: *What factors affect worker perception of pay fairness on microtask crowdsourcing platforms?* Motivated by the ethical paradox of the gig economy, where work autonomy and flexibility are often associated with wage suppression and lack of worker protection, this study employed a hybrid methodological framework to bridge the gap between qualitative narratives of work experiences and quantitative assessments of pay fairness.

In conclusion, this thesis argues that fairness in microtask crowdsourcing is not simply an objective measure of dollars per hour, but a subjective socio-technical evaluation of how workers' time, effort, and dignity are recognized and respected in digital labor exchange. Drawing on workers' voices through computational text analysis, this study provides both a theoretical lens and a methodological approach for advancing a more equitable future for platform-enabled on-demand labor.

More broadly, this research contributes to the growing scholarship on digital labor platforms by demonstrating that organizational justice theory, when adapted to the distinctive features of algorithmically mediated platform work, offers a foundational framework for understanding fairness in nonstandard employment. As platform-enabled labor continues to expand across sectors and geographies, the hybrid methodology developed in this study, integrating unsupervised text mining, theory-driven interpretation, and predictive modeling, serves as a scalable approach for examining worker experiences on a wider range of digital labor platforms. In this way, the study informs ongoing efforts by researchers, platform designers, and policymakers to promote greater transparency, strengthen accountability, and foster more equitable working conditions in the evolving platform-enabled gig economy.

7.2 Limitations and Future Work

This study has several limitations that suggest directions for future research. First, it focuses on microtask crowdsourcing, a specific type of platform work characterized by short, bite-sized tasks compensated with small payments. In other types of crowdsourcing, such as project-based freelancing, factors pertaining to the nature of tasks (e.g., easy earnings or time efficiency) may not emerge as salient aspects of workers' positive experiences. Furthermore, although MTurk serves as a prominent empirical setting for microtask crowdsourcing, it is not fully representative of the governance practices of other microtask crowdsourcing platforms. For example, Prolific enforces minimum wage requirements and exercises stronger governance over payment conditions. As a result, the findings may not be fully generalizable to microtask contexts marked by more robust platform protections. Future research could extend and validate the findings across digital labor platforms that vary in task types and governance arrangements.

Second, the dataset is subject to self-selection bias in two aspects. On the one hand, the dataset captures only workers who voluntarily choose to post reviews on TurkerView, and these workers may differ from the broader MTurk workforce in terms of engagement, experience, or propensity to express strong opinions. On the other hand, workers may be especially motivated to review tasks that provoke particularly strong positive or negative reactions, which might lead to the underrepresentation of more moderate or neutral experiences. These self-selection effects may amplify the apparent salience of extreme fairness perceptions relative to their prevalence in the broader worker population. Future research could address this limitation by triangulating TurkerView data with other data sources, such as platform-administered surveys, interviews, or experimental designs, to assess the representativeness of the review-based findings.

Third, the predictive models do not include potential confounding factors such as worker demographics (e.g., age, gender, education, geographic location) and more detailed task characteristics (e.g., task category, task batch size). These omissions stem largely from the constraints of web scraping and data availability. These variables may moderate the relationships between topic factors and pay fairness perceptions. For example, workers in lower-income regions may weigh distributive justice concerns more heavily, while

workers with higher educational attainment may be more sensitive to procedural fairness. Future research could incorporate these variables, potentially through linked datasets or platform APIs that provide richer metadata, to develop a more comprehensive model of the determinants of pay fairness perception.

Finally, this study has yet to examine the interplay between perceived fairness and normative fairness, which refers to established principles or standards that define what should be considered fair (Liu & Mendez, 2022) and are often grounded in ethical theories and social norms. Whereas perceived fairness reflects workers' subjective judgments, which may be formed through cognitive heuristics, reference-point comparisons, and bounded rationality (Kahneman & Tversky, 1979; Konow, 2003), it may also be influenced by broader social norms of fairness. As crowdsourcing fosters nonstandard forms of employment, future research could explore how microworkers develop compensation norms in online labor markets and how these emerging norms, in turn, shape their perceptions of pay fairness.

References

Adams, J. S. (1965). Inequity In social exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 267–299). Academic Press. https://doi.org/10.1016/S0065-2601(08)60108-2

Allport, G. W. (1954). *The nature of prejudice* (Abridged, [Nachdr.]). Doubleday.

Alpar, P., & Osterbrink, L. (2020). Antecedents and consequences of perceived fairness in pay for crowdwork. *Proceedings of the 41st International Conference on Information Systems*, 1–17.

Amazon Mechanical Turk. (2024). https://www.mturk.com/

Angelov, D. (2020). *Top2Vec: Distributed representations of topics* (arXiv:2008.09470). arXiv. https://doi.org/10.48550/arXiv.2008.09470

Arun, R., Suresh, V., Veni Madhavan, C. E., & Narasimha Murthy, M. N. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 391–402). Springer. https://doi.org/10.1007/978-3-642-13657-3_43

Barends, A. J., & de Vries, R. E. (2019). Noncompliant responding: Comparing exclusion criteria in MTurk personality research to improve data quality. *Personality and Individual Differences*, *143*, 84–89. https://doi.org/10.1016/j.paid.2019.02.015

Behl, A., Sheorey, P., Jain, K., Chavan, M., Jajodia, I., & Zhang, Z. (Justin). (2021). Gamifying the gig: Transitioning the dark side to bright side of online engagement. *Australasian Journal of Information Systems*, *25*, 1–34. https://doi.org/10.3127/ajis.v25i0.2979

Berente, N., Seidel, S., & Safadi, H. (2019). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50–64. https://doi.org/10.1287/isre.2018.0774

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 Organization. Politics Collection (2542200794).

https://ezproxy.library.yorku.ca/login?url=https://www.proquest.com/reports/digital-labour-platforms-future-work-towards/docview/2542200794/se-2?accountid=15182

Beugre, C. D., & Baron, R. A. (2001). Perceptions of Systemic Justice: The Effects of Distributive, Procedural, and Interactional Justice. *Journal of Applied Social Psychology*, *31*(2), 324–339. https://doi.org/10.1111/j.1559-1816.2001.tb00199.x

Bies, R. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations*, *1*, 43–55.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2002). Latent dirichlet allocation. *Journal of Machine Learning Research*, 601–608.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York city cabdrivers: One day at a time*. *The Quarterly Journal of Economics*, *112*(2), 407–441. https://doi.org/10.1162/003355397555244

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, Advances in Machine Learning and Computational Intelligence*, *72*(7), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior and Organization*, *90*(Complete), 123–133. https://doi.org/10.1016/j.jebo.2013.03.003

Chen, D. L., & Horton, J. J. (2016). Research Note—Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts. *Information Systems Research*, *27*(2), 403–423. https://doi.org/10.1287/isre.2016.0633

Chen, M. K. (2016). *Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform*. 455–455. https://doi.org/10.1145/2940716.2940798

Chen, W.-C., Suri, S., & Gray, M. L. (2019). More than money: Correlation among worker demographics, motivations, and participation in online labor market. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 134–145. https://doi.org/10.1609/icwsm.v13i01.3216

Chiu, H., Palupi, G., & Zhu, Y.-Q. (2022). Workers' Affective Commitment in The Gig Economy: The Role of IS Quality, Organizational Support, and Fairness. *Pacific Asia Journal of the Association for Information Systems*, *14*(3). https://doi.org/10.17705/1pais.14303

Cobanoglu, C., Cavusoglu, M., & Turktarhan, G. (2021). A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of Amazon mechanical turk (MTurk). *Journal of Global Business Insights*, *6*(1), 92–97. https://doi.org/10.5038/2640-6489.6.1.1177

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*, *98*(2), 199–236. https://doi.org/10.1037/a0031757

Cropanzano, R., Bowen, D. E., & Gilliland, S. W. (2007). The management of organizational justice. *Academy of Management Perspectives*, *21*(4), 34–48. https://doi.org/10.5465/amp.2007.27895338

D'Cruz, P., & Noronha, E. (2016). Positives outweighing negatives: The experiences of indian crowdsourced workers. *Work Organisation, Labour & Globalisation*, *10*(1), 44–63. https://doi.org/10.13169/workorgalaboglob.10.1.0044

Deng, X., & Joshi, K. D. (2016). Why Individuals Participate in Micro-task Crowdsourcing Work Environment: Revealing Crowdworkers' Perceptions. *Journal of the Association for Information Systems*, *17*(10), 648–673. https://doi.org/10.17705/1jais.00441

Deng, X., Joshi, K. D., & Galliers, R. D. (2016). The Duality of Empowerment and Marginalization in Microtask Crowdsourcing: Giving Voice to the Less Powerful Through Value Sensitive Design. *MIS Quarterly*, *40*(2), 279–302.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Di Gangi, P., Howard, J., McAllister, C., & Thatcher, J. (2024). The influence of political skill and community capabilities on microtask worker hourly wage: A mixed methods study of mechanical turk. *Journal of the Association for Information Systems*, *25*(4), 890–935. https://doi.org/10.17705/1jais.00858

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, 135–143. https://doi.org/10.1145/3159652.3159661

Dornstein, M. (1989). The fairness judgements of received pay and their determinants. *Journal of Occupational Psychology*, *62*(4), 287–299. https://doi.org/10.1111/j.2044-8325.1989.tb00501.x

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal*, *30*(1), 114–132. https://doi.org/10.1111/1748-8583.12258

Dupuis, M., Renaud, K., & Searle, R. (2022). Crowdsourcing quality concerns: An examination of Amazon's mechanical turk. *Proceedings of the 23rd Annual Conference on Information Technology Education, SIGITE '22*, 127–129. https://doi.org/10.1145/3537674.3555783

Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

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*(4), 987–1005. https://doi.org/10.1007/s10551-017-3607-2

Fowler, C., Jiao, J., & Pitts, M. (2023). Frustration and ennui among Amazon MTurk workers. *Behavior Research Methods*, *55*(6), 3009–3025. https://doi.org/10.3758/s13428-022-01955-9

Gandini, A. (2019). Labour process theory and the gig economy. *Human Relations*, *72*(6), 1039–1056. https://doi.org/10.1177/0018726718790002

Gillies, M., Murthy, D., Brenton, H., & Olaniyan, R. (2022). *Theme and topic: How qualitative research and topic modeling can Be brought together* (arXiv:2210.00707). arXiv. https://doi.org/10.48550/arXiv.2210.00707

Graham, M., & Woodcock, J. (2018). Towards a fairer platform economy: Introducing the fairwork foundation. *Alternate Routes*, *29*, 242–253.

Greenberg, J. (1990). Organizational Justice: Yesterday, Today, and Tomorrow. *Journal of Management*, *16*(2), 399–432. https://doi.org/10.1177/014920639001600208

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, *380*(6650), 1108–1109. https://doi.org/10.1126/science.adi1778

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, *13*(2), 586–632. (137561635). https://doi.org/10.5465/annals.2017.0099

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. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3174023

Hornuf, L., & Vrankar, D. (2022). Hourly Wages in Crowdworking: A Meta-Analysis. *Business & Information Systems Engineering*, *64*(5), 553–573. https://doi.org/10.1007/s12599-022-00769-5

Horton, J. J., & Chilton, L. B. (2010). The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM Conference on Electronic Commerce, EC '10*, 209–218. https://doi.org/10.1145/1807342.1807376

Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, 611–620. https://doi.org/10.1145/2470654.2470742

Irani, L. C., & Silberman, M. S. (2016). Stories We Tell About Labor: Turkopticon and the Trouble with "Design." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, 4573–4586. https://doi.org/10.1145/2858036.2858592

Jabagi, N., Croteau, A.-M., Audebrand, L. K., & Marsan, J. (2025). Do algorithms play fair? Analysing the perceived fairness of HR-decisions made by algorithms and their impacts on gig-workers. *The International Journal of Human Resource Management*, *36*(2), 235–274. https://doi.org/10.1080/09585192.2024.2441448

Jiang, L., & Wagner, C. (2024). How low is low? Crowdworker perceptions of microtask payments in work versus leisure situations. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24*, 1–11. https://doi.org/10.1145/3613904.3642601

Jiang, L., Wagner, C., & Nardi, B. (2015). Not Just in it for the Money: A Qualitative Investigation of Workers' Perceived Benefits of Micro-task Crowdsourcing. *2015 48th Hawaii International Conference on System Sciences*, 773–782. https://doi.org/10.1109/HICSS.2015.98

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kim, T.-Y., Wang, J., Chen, T., Zhu, Y., & Sun, R. (2019). Equal or equitable pay? Individual differences in pay fairness perceptions. *Human Resource Management*, *58*(2), 169–186. https://doi.org/10.1002/hrm.21944

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13*, 1301–1318. https://doi.org/10.1145/2441776.2441923

Kjell, O. N. E., Kjell, K., Garcia, D., & Sikström, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods*, *24*(1), 92–115. https://doi.org/10.1037/met0000191

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods*, *21*(3), 733–765. https://doi.org/10.1177/1094428117722619

Konow, J. (2003). Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature*, *41*(4), 1188–1239. https://doi.org/10.1257/002205103771800013

Kummerfeld, J. K. (2021). Quantifying and Avoiding Unfair Qualification Labour in Crowdsourcing. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 343–349). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-short.44

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C.

(2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

Lee, S., Hong, S., Shin, W.-Y., & Lee, B. G. (2023). The experiences of layoff survivors: Navigating organizational justice in times of crisis. *Sustainability*, *15*(24), 16717. https://doi.org/10.3390/su152416717

Leventhal, G. S. (1980). What should Be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social Exchange: Advances in Theory and Research* (pp. 27–55). Springer US. https://doi.org/10.1007/978-1-4613-3087-5_2

Liu, J. T., & Mendez, M. J. (2022). The Attributional–Counterfactual Theory of Need: Integrating Theories to Predict Need Norm Use. *Business Ethics Quarterly*, *32*(1), 103–135. https://doi.org/10.1017/beq.2021.1

Marshall, C. C., Goguladinne, P. S. R., Maheshwari, M., Sathe, A., & Shipman, F. M. (2023). Who broke amazon mechanical turk? An analysis of crowdsourcing data quality over time. *Proceedings of the 15th ACM Web Science Conference 2023, WebSci '23*, 335–345. https://doi.org/10.1145/3578503.3583622

Mason, W., & Watts, D. J. (2010). Financial incentives and the "performance of crowds." *ACM SIGKDD Explorations Newsletter*, *11*(2), 100–108. https://doi.org/10.1145/1809400.1809422

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2023). Is it ethical to use Mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages. *Behavior Research Methods*, *55*(8), 4048–4067. https://doi.org/10.3758/s13428-022-02005-0

Namvar, M., Akhlaghpour, S., Boyce, J., & Khajedehi, S. S. (2022). Iterative seed word generation for interactive topic modelling: A mixed text processing and qualitative content analysis approach. *ICIS 2022 Proceedings*. https://aisel.aisnet.org/icis2022/online_reviews/online_reviews/6

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3–42. https://doi.org/10.1177/0049124117729703

Odacioglu, E. C., & Zhang, L. (2022). *Text mining for rendering theory: Integrating topic modeling to grounded theory* (SSRN Scholarly Paper No. 4141372). Social Science Research Network. https://doi.org/10.2139/ssrn.4141372

Pittman, M., & Sheehan, K. (2016). Amazon's Mechanical Turk a Digital Sweatshop? Transparency and Accountability in Crowdsourced Online Research. *Journal of Media Ethics*, *31*(4), 260–262. https://doi.org/10.1080/23736992.2016.1228811

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*, 399–408. https://doi.org/10.1145/2684822.2685324

Saito, S., Chiang, C.-W., Savage, S., Nakano, T., Kobayashi, T., & Bigham, J. (2019). TurkScanner: Predicting the Hourly Wage of Microtasks. *The World Wide Web Conference*, 3187–3193. https://doi.org/10.1145/3308558.3313716

Salminen, J., Kamel, A. M. S., Jung, S.-G., Mustak, M., & Jansen, B. J. (2023). Fair compensation of crowdsourcing work: The problem of flat rates. *Behaviour & Information Technology*, *42*(16), 2871–2892. https://doi.org/10.1080/0144929X.2022.2150564

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter* (arXiv:1910.01108). arXiv. https://doi.org/10.48550/arXiv.1910.01108

Savage, S., Chiang, C. W., Saito, S., Toxtli, C., & Bigham, J. (2020). Becoming the Super Turker:Increasing Wages via a Strategy from High Earning Workers. *Proceedings of The Web Conference 2020, WWW '20*, 1241–1252. https://doi.org/10.1145/3366423.3380200

Scheel, T. E., Mendling, J., & Schlagwein, D. (2025). Task performance and platform commitment: The role of job design and worker motivation in microtasking crowdsourcing. *Frontiers in Organizational Psychology*, *3*. https://doi.org/10.3389/forgp.2025.1678801

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The Polarity of Online Reviews: Prevalence, Drivers and Implications. *Journal of Marketing Research*, *57*(5), 853–877. https://doi.org/10.1177/0022243720941832

Schulze, L., Trenz, M., Cai, Z., & Tan, C.-W. (2023). *Fairness in algorithmic management: How practices promote fairness and redress unfairness on digital labor platforms*. 196–205. https://doi.org/10.24251/HICSS.2023.024

Shi, Z. (Michael), Lee, G. M., & Whinston, A. B. (2016). Toward a better measure of business proximity: Topic modeling for industry Intelligence1. *MIS Quarterly*, *40*(4), 1035–1056. https://doi.org/10.25300/MISQ/2016/40.4.11

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, *82*(3), 434–443. https://doi.org/10.1037/0021-9010.82.3.434

Song, X., Lowman, G. H., & Harms, P. (2020). Justice for the crowd: Organizational justice and turnover in crowd-based labor. *Administrative Sciences*, *10*(4), 93. https://doi.org/10.3390/admsci10040093

Tan, Z. M., Aggarwal, N., Cowls, J., Morley, J., Taddeo, M., & Floridi, L. (2021). The ethical debate about the gig economy: A review and critical analysis. *Technology in Society*, *65*, 101594. https://doi.org/10.1016/j.techsoc.2021.101594

Toxtli, C., Suri, S., & Savage, S. (2021). Quantifying the invisible labor in crowd work. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–26. https://doi.org/10.1145/3476060

Vakulenko, S., Müller, O., & Brocke, J. (2014). Enriching iTunes app store categories via topic modeling. *ICIS 2014 Proceedings*. https://aisel.aisnet.org/icis2014/proceedings/EBusiness/36

Vallas, S., & Schor, J. B. (2020). What do platforms do? Understanding the gig economy. *Annual Review of Sociology*, *46*(Volume 46, 2020), 273–294. https://doi.org/10.1146/annurev-soc-121919-054857

Whiting, M. E., Hugh, G., & Bernstein, M. S. (2019). Fair Work: Crowd Work Minimum Wage with One Line of Code. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *7*, 197–206. https://doi.org/10.1609/hcomp.v7i1.5283

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. https://doi.org/10.1177/0950017018785616

Wood, A. J., Lehdonvirta, V., & Graham, M. (2018). Workers of the Internet unite? Online freelancer organisation among remote gig economy workers in six Asian and African countries. *New Technology, Work and Employment*, *33*(2), 95–112. https://doi.org/10.1111/ntwe.12112

Yang, J., Valk, C. van der, Hoßfeld, T., Redi, J., & Bozzon, A. (2018). How do crowdworker communities and microtask markets influence each other? A data-driven study on Amazon mechanical turk. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *6*, 193–202. https://doi.org/10.1609/hcomp.v6i1.13335

Yang, Y., & Subramanyam, R. (2023). Extracting Actionable Insights from Text Data: A Stable Topic Model Approach: MIS Quarterly. *MIS Quarterly*, *47*(3), 923–954. https://doi.org/10.25300/MISQ/2022/16957

Ye, T., You, S., & Robert Jr., L. (2017). When Does More Money Work? Examining the Role of Perceived Fairness in Pay on the Performance Quality of Crowdworkers. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), 327–336. https://doi.org/10.1609/icwsm.v11i1.14876

Zhang, D., Zhang, K., Yang, Y., & Schweidel, D. A. (2024). Tm-okc: An unsupervised topic model for text in online knowledge communities. *MIS Quarterly*, *48*(3), 931–978. (179387169). https://doi.org/10.25300/MISQ/2023/17885

Appendices

Appendix A: Representative Quotes for Topics Emerging from Positive Reviews

Topic 1: Generous Pay

- *"Excellent overall pay for this study; you see lots of HITs that have a high hourly rate and quite a few that have a high total dollar value, but rarely one that has both."* (Review #1301658)
- *"Fair pay rating and a good chunk of change to make this very worth it."* (Review #1554553)
- *"If you perform the task properly, the pay rating is very good. The task itself is fairly engaging as well. Approved and bonus paid within hours."* (Review #686499)
- *"Clear instructions, exemplary survey, prompt approval and good pay rate. This requester is fair and trustworthy."* (Review #1718207)
- *"Excellent requester. Zero tech issues on well-designed and engaging HITs, generous pay rating, and quick approvals."* (Review #1687194)

Topic 2: Bonus Opportunities

- *"Screener at the start. If you are selected, can continue with the rest of the survey. They state you will receive a $1.50 base payment, so assuming this will be bonused at $1.35. There is an additional opportunity for bonus, but this is based on a lottery. Edit: received the promised $1.35 bonus to bring the base payment to $1.50."* (Review #1446877)
- *"Potential bonus in 2 parts: one out of every ten participants who complete this survey will have their choices implemented for real. Since these involve winning between $20 and $60, this means that you might earn a significant amount of extra money. For each correct guess you make, you will earn $0.25. Paid within 1 week."* (Review #1158753)
- *"Extra bonus up to $1 is offered; while not a totally random drawn bonus, it is very much based on probability. I would guess most participants will not receive much, if any, bonus on this task. Edited to add: I did receive a $0.60 bonus on this task."* (Review #922714)

- *"Three tasks for bonus. Bonus opportunity, but YMMV as there is some randomness built in. Edit: received a $4.92 bonus same day. Appears that money earned in round 1 was also added."* (Review #1391115)
- *"Bonus potential: You will be paid a $2 base pay PLUS $1 for each of the four comprehension questions you answer correctly PLUS one of your choices in the main task will be randomly chosen for payment. Your additional randomly chosen bonus can earn you up to an additional $5."* (Review #1161402)

Topic 3: Easy Earnings

- *"Pretty easy task once you understand it. The tasks are pretty straightforward and involve little writing."* (Review #949881)
- *"Straightforward and easy. I always tend to take more time than necessary so quicker for others. No writing required."* (Review #805650)
- *"Some writing, but nothing too bad. Overall a pretty quick study. Don't pass this one up."* (Review #1713623)
- *"Super fast, easy and kind of fun. Potential raffle prize."* (Review #698122)
- *"Fast and easy. Small amount of writing. Could be done a little bit faster."* (Review #1237164)

Topic 4: Task Instructional Clarity

- *"Answer some societal type questions - read 5 different questions/scenarios answer each one using sliders - describe the questions - answer the same societal type questions as before - read about a type of product - answer some questions in relation to this product - answer some economic type questions - demographics."* (Review #1541120)
- *"Read about someone on social media - answer various questions about this person - answer questions about brands and products - read a statement - answer one question - read about 5 different products - answer various questions on each product - short demographics."* (Review #1480104)
- *"Answer a few questions about a type of product - read some information about this product - answer 2 questions - read about a scenario - answer one question - write as instructed - answer various questions - read about another scenario - answer 3 questions - answer self/technology type questions - demographics."* (Review #1559626)

- *"Answer some political demographic questions - read a statement - answer some questions in relation to what you have just read - read another statement - answer questions in relation to that statement - answer various societal type questions - demographics."* (Review #1548179)
- *"Demographics - read instructions - answer various self type questions - assess the results - answer questions in relation to the results - explain your thoughts no limit in character or words - answer a few more questions - more demographics."* (Review #1494645)

Topic 5: Time Efficiency

- *"Each block should take no more than 13 minutes (multiply by 6 blocks). Initial setup is no more than 5 minutes. Upload at the end took 4 minutes. Timer is sufficiently long."* (Review #882337)
- *"Took way less time than the estimated 60 minutes! Glad I took a gamble on this one."* (Review #745870)
- *"Long timer. Grab a coffee and take advantage of that timer and take a few breaks."* (Review #1177781)
- *"I didn't think this was nearly as long or gruesome as the other reviews made it sound. You can save your progress and come back. There's no penalty for wrong answers. No time limits. The HIT itself had a 3-day timer, so you've got plenty of time to pace yourself."* (Review #1234260)
- *"Short task, most of the time is spent waiting (for good reason). The forced break gave me enough time to run and refill my coffee in the kitchen and make it back just before the timer ran out. Possibility for future surveys."* (Review #205931)

Topic 6: Requester Responsiveness

- *"He was very communicative. I had emailed and he responded telling me what I had done wrong. It was obvious so I don't know how I missed it in the first place but between the people who missed the control questions and the people who were rejected because they did more than one of his hits they decided to reverse all the rejections."* (Review #51969)
- *"I wrote a polite note asking what I did wrong. He responded with a detailed message and I was incorrect. Hours later he messaged me and said the rejection would be reversed. I did not ask, I was wrong. He said it was confusing to a lot of people. Stand up guy."* (Review #51921)

- *"Heard back from the requester and they made an error and rejected everyone accidentally. Said that they have a call scheduled with Amazon to try to fix the problem. Rejection overturned!"* (Review #1182216)
- *"Initially this was a rejection. I was rejected because my completion code did not match the code I put in. I explained the situation and sent a copy of my Prolific account showing my Prolific ID number. She reversed the rejection, for which I am grateful."* (Review #1305058)
- *"She set a new record this afternoon on quick response times when an MTurk participant has a problem with a survey. Her HITs are nothing out of the ordinary, but she not only answers you if there's a problem, but she tends to answer you FAST. I reported an issue with a Qualtrics error today, and within around three minutes she had approved the HIT submission."* (Review #1544826)

Topic 7: Well-structured Task

- *"Short hypothetical that develops over a couple of pages with a handful of questions and a required written response mixed in. Demographics at the end."* (Review #1753802)
- *"Quick. Brief hypothetical, followed by a handful of questions. Short demographics at the end."* (Review #1323013)
- *"Quick. Short hypothetical followed by a couple of pages of light bubbles. Brief demographics at the end."* (Review #1210964)
- *"Quick. 2 side-by-side comparison of images, each followed by a short writing prompt. A page with about 5 bubbles. Short demographics at the end."* (Review #1411803)
- *"Listen to 60 music samples approximately 10 seconds in length, then select whether you recognize the song, and when you believe you first heard it. That's it!"* (Review #1622007)

Appendix B: Representative Quotes for Topics Emerging from Negative Reviews

Topic 1: Task Interface Friction

- *"Steaming pile of garbage. I would not recommend doing this HIT. I feel like I complete HITs pretty fast on average and this still only came out to $5 per hour. On top of that it was slider hell with an extremely inaccurate progress bar."* (Review #4488)

- *"Straight up garbage. No progress bar. Repeats the same questions over and over again. Very repetitive. Sliders UI sometimes loads slowly and timed sections make you stay on a page way longer than you need to be."* (Review #836949)
- *"Extremely tedious and repetitive; excruciatingly so. So, so underpaid. Only do this hit if you are extremely bored and have nothing else to do, or it's a slow Sunday, or you hate yourself and like pain."* (Review #782072)
- *"Horrible HIT with horrible interface. Entering text describing pictures that cannot be edited except with the backspace key. Badly, badly, badly paid."* (Review #961071)
- *"God awful face rating task that drags on forever. You see the same faces over and over again and rate them one at a time on some kind of scale. Bubbles, sliders, and you will quickly grow to dislike all of the faces."* (Review #144595)

Topic 2: Excessive Task Length

- *"Repetitive questions. Asks several questions over and over again. Will ask the same five questions and then ask them again with slightly different phrasing. Writing is involved. They ask you to write out a brief passage of thoughts."* (Review #855333)
- *"Excessive amount of writing. About a dozen writing questions and a lot of bubble questions."* (Review #949221)
- *"Almost all writing. 8 open-writing questions with minimum character requirements. Underpaid for the amount of writing and the detail of writing they want."* (Review #1219873)
- *"2 experiential writing prompts followed by tons of personality bubbles and product evaluations."* (Review #1155490)
- *"A lot of reading and writing, followed by tons of bubble questions."* (Review #692776)

Topic 3: Technical Delays and Lag

- *"Each audio clip takes 10-15 seconds to load on my otherwise fast connection. There were 80 clips in this HIT, so that's a lot of wasted loading time. Pay rate would be low even without the extra delays."* (Review #271025)

- *"Their server was running slowly for most of the HIT. I refreshed once to try to get the survey to move and it kept my progress. I tried it again on the second task and it reset my progress."* (Review #1708874)
- *"Every page is on a timer. You'll spend half your time waiting for the 'next' arrow to appear."* (Review #505910)
- *"Parts of this moves at its own pace. Time spent waiting for main task to load. Main task forces you to fullscreen."* (Review #1450640)
- *"Had to watch 20 short videos that loaded slowly and kept buffering. A 30 second video sometimes took 2 minutes to play. Would have been a decent hit without this problem."* (Review #210069)

Topic 4: Task/Survey System Errors

- *"Broken code link at the end. To fix it, take the URL starting with HTTP and ending with Dynamic Key. Toss that URL into a new browser window. Now, go back to the HIT window and right click the survey link and copy link address. Enter that link address into that new browser window where it tells you."* (Review #694002)
- *"Asks for Prolific ID at end and redirects to Prolific. I gave my mTurk ID and copied what I assume was the code from the URL."* (Review #1117698)
- *"Two separate completion codes were provided, the Sawtooth link told me to paste the code into Mturk, as well as the Qualtrics survey had me paste the Sawtooth code into it and then that gave me a code for mturk. I submitted with the code from the Qualtrics."* (Review #1677917)
- *"Error Message on last page of survey instead of providing a completion code: 'Deprecated - Please Use' followed by a Qualtrics support URL for web service random number generation."* (Review #1382028)
- *"Broken survey. Keeps redirecting to additional surveys instead of ending with a code. I completed three surveys before noping out and adding an explanation in the code box."* (Review #1464366)

Topic 5: Attention/Comprehension Check Failures

- *"Failed the comprehension for the instructions check twice at 4/5 despite changing only one answer. I'm pretty sure they have it set up so you auto fail regardless of your answers and make you go through the instructions more than once."* (Review #101844)

- *"My submission was rejected for 'failed attention check'. I was very careful while completing the study and I'm 100% sure I did not miss any attention checks. I think this is just an attempt by the requester to not pay."* (Review #835607)
- *"Said I failed the attention check even though I know for certain I passed the attention check."* (Review #1255498)
- *"Rejected my hit claiming I missed one of two attention checks. I retook the survey and only one attention check is present."* (Review #1101816)
- *"Attention check states to select disagree. There are two options with disagree in that question."* (Review #537373)

Topic 6: Uncertainty in Earnings

- *"This HIT is absolute garbage. What the HIT does not tell you is that in 5/10 rounds your random valuation makes it so that your max earnings for that round are 0. So when they tell you one round will randomly be chosen for your bonus at the end, what they really mean is you have a 50% chance of a bonus and a 50% chance of being slave labor for their experiment."* (Review #136169)
- *"Promises of additional bonuses are exaggerated. They say each task has an equal chance of being chosen. Unfortunately, a good portion of the tasks have 0 or negative bonuses attached to them. Without additional bonuses this task pays below minimum wage and is not worth the hassle."* (Review #430099)
- *"They tell you at the end of the experiment that 20% of participants will actually get the bonus from the experiment. 20% chance of generous and 80% chance of low pay."* (Review #154083)
- *"Additional bonus can be as low as 50 cents depending on luck. You could also end up with no bonus at all. Pay is just okay without the bonus."* (Review #604632)
- *"Low pay. Bonus is determined by a random chance, and the amount is determined by another random chance, so don't count on anything."* (Review #249693)

Topic 7: Nonresponsive or Erroneous Rejections

- *"Auto rejected immediately. I have disputed the rejection and will update if I hear anything. After looking at the other reviewers comments and then realizing this requestor only has put out 1 HIT so far and auto rejected*

anyone who took it, I think the requestor either really doesn't know what they are doing, or is data stealing. EDIT: The rejection was overturned. I received a very polite and apologetic email from them concerning the incident, it was a mistake." (Review #1155191)

- *"My work was rejected with the generic 'your work did not meet our standards' message. I contacted them twice in an attempt to find out why. I have received no reply 2 weeks later. 3 weeks later, 3 messages later, still no reply. This requester's communication has been absolutely unacceptable, or rather completely non-existent. Avoid."* (Review #563473)
- *"I and several other turkers received a rejection stating simply 'Unable to validate'. I know they've received multiple requests to further explain the rejection and have yet to hear about anyone getting a response. UPDATE: Got an email stating they reversed the mass rejections, putting it down to an error where all the data was overwritten somehow, along with an invitation to retake the survey."* (Review #721977)
- *"I've done hundreds if not thousands of these with no issues but recently they have gone on a rejection spree and have given me 6 rejections in a row with no explanation. When I contact them they just say my answer wasn't a quality answer. Which makes no sense how can 1000 over a month be ok and 6 out of 20 be bad in the same day."* (Review #1294389)
- *"Rejected along with everyone else it seems. New requester, may not understand the ethics of paying people for their work or how the platform works entirely. I'll cross my fingers they see the multiple rejection disputes and do the right thing."* (Review #1505709)