The Making of Performative Accuracy in AI Training: Precision Labor and Its Consequences

Ben Zefeng Zhang
University of Michigan
Ann Arbor, Michigan, USA
bzfzhang@umich.edu

Tianling Yang
Technische Universität Berlin
Berlin, Germany
Weizenbaum Institut
Berlin, Germany
tianling.yang@tu-berlin.de

Milagros Miceli
DAIR Institute
Berlin, Germany
Technische Universität Berlin
Berlin, Germany
Weizenbaum Institut
Berlin, Germany
m.miceli@tu-berlin.de

Oliver L. Haimson
University of Michigan
Ann Arbor, Michigan, USA
haimson@umich.edu

Michaelanne Thomas
University of Michigan
Ann Arbor, Michigan, USA
mmtd@umich.edu

Abstract

Accuracy and precision are central values in the AI communities and the technology sector. This paper provides empirical evidence on the construction and organizational management of technical accuracy, demonstrating how technology companies' preoccupation with such values leads to harm. Drawing on nine months of multi-sited ethnographic fieldwork in China, we document how AI trainers' everyday work practices, challenges, and harms stem from clients' demands for high levels of technical accuracy. We introduce the concept of *precision labor* to unpack the labor dimension of constructing and performing accuracy in AI training. This concept highlights the hidden and excessive labor required to reconcile the ambiguity and uncertainty involved in this process. We argue that *precision labor* offers a new lens to illuminate three critical aspects of AI training: 1) the negative health and financial impacts of hidden and excessive labor on AI workers; 2) emerging harms, including workers' subordinate roles to machines and financial precarity; and 3) a conceptual contribution to contexts beyond AI training. This contribution re-centers arbitrariness in technical production, highlights the excessive demands of precision labor, and examines the legitimization of labor and harm. Our study also contributes to existing scholarship on the prevailing values and invisible labor in AI production, underscoring accuracy as performative rather than self-evident and unambiguous. A *precision labor* lens challenges the legitimacy and sustainability of relentlessly pursuing technical accuracy, raising new questions about its consequences and ethical implications. We conclude by proposing recommendations and alternative approaches to enhance worker agency and well-being.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713112

Keywords

data work, AI training, digital labor, microwork, precision labor, ethnography, China

ACM Reference Format:
Ben Zefeng Zhang, Tianling Yang, Milagros Miceli, Oliver L. Haimson, and Michaelanne Thomas. 2025. The Making of Performative Accuracy in AI Training: Precision Labor and Its Consequences. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713112

1 Introduction

> *[Precision] connotes trustworthiness and elegance in the actions or products of humans and machines. Precision is everything that ambiguity, uncertainty, messiness, and unreliability are not. It is responsible, nonemotional, objective, and scientific. It shows quality. These "values of precision" have become part of our heritage.*
> — M. Norton Wise, *The Values of Precision*

Accuracy and precision carry immense weight in the development and deployment of modern technology. These concepts often starkly contrast with ambiguity, uncertainty, and messiness [98]. They are closely associated with objectivity [51], scalability [87], and intelligence [13], thereby serving as key tenets that legitimize and monetize technology products. Despite their ubiquity, accuracy and precision are difficult to pin down with unifying definitions. Drawing on prior works [98], we propose that precision encompasses accuracy, and in contexts such as AI training, they can be used interchangeably. As prevailing goals and values in commercial technology production, a relentless pursuit of technical accuracy[1] can pressure and harm workers. What does it take to achieve a high level of technical accuracy? What are the harms resulting from technology companies' excessive focus on technical accuracy and precision, and who incurs the greatest burdens in this process?

[1] We refer to *accuracy* as one of the central values in ML communities and use *technical accuracy* to emphasize technical arrangements around accuracy on the ground, such as strict accuracy standards, and *performative accuracy* to underscore the excessive and performative aspects of labor involved in achieving arbitrary accuracy standards.

This paper documents the construction and management of accuracy and precision in AI training. This documentation uncovers the emergence of new harms, excessive labor, and hidden labor in AI training. We provide an ethnographic account of AI trainers[2], a relatively understudied but emergent category of laborers in China — a pivotal site in global AI production [43].

AI training in China highlights the gaps and essential role of human labor in enabling, facilitating, and sustaining the development and maintenance of AI systems. In response to these gaps, "AI trainer" has been established as an occupation. We use the term to refer to AI workers who perform critical tasks behind the scenes in the AI industry, including data collection and annotation, AI-enabled system testing, operation, implementation and maintenance, and training of other AI trainers. As this occupation continues to evolve, its professional boundaries and characteristics remain in flux. In its current stage, we observed that AI trainers' work largely overlaps with data work [56] that produces machine learning (ML) [3]datasets. These tasks involve data generation, annotation, and output evaluation and verification. However, the role of AI trainers extends beyond that of traditional data workers [56] in the AI data pipeline [60]. AI trainers work in different phases of AI testing, implementation, and maintenance, which is part of uncredited but essential "ghost work" [25]. Unlike ghost work, which remains largely invisible within the AI industry, the professionalization of AI trainers as a stable tech workforce promises workers greater visibility and recognition. Establishing AI trainers as a formal occupation goes beyond merely elevating the status of data workers; it is part of an ongoing institutionalization process. This process involves multiple actors, informal rules, formal rules, and structures that integrate AI trainers into the fabric of society. It includes formalizing and standardizing the skills and expertise required by technology companies, offering training and certification through authorized agencies, and creating mechanisms for social mobility.[4] These institutional arrangements could offer AI trainers material benefits, such as higher income levels tied to occupational titles and ranks, which already distinguish AI trainers in China from those in other geographical contexts.

Previous literature has examined the hidden labor behind AI technologies [20, 25]. With a specific focus on the human labor required to produce ML datasets, existing work has investigated work practices, workflows, and social and organizational contexts [56, 57, 63, 67, 93, 101]. However, less attention has been given to the construction and management of technical accuracy from an organizational perspective. Labor in data production is often outsourced through crowdsourcing platforms [73, 90] and business process outsourcing companies (hereafter BPOs) [57, 93]. With diverse institutional arrangements [60], AI data annotation centers in China and their hybrid management reflect the particularities of the Chinese context, which we will elaborate on later in this paper.

By investigating accuracy and precision in commercial AI training, we show how these values are perceived, established, and managed on the ground. We found that the high and sometimes arbitrary accuracy standards set by the clients from technology companies and managers demand excessive and hidden labor, which harms the AI workers involved.

These harms include, but are not limited to, excessive training and revisions to meet extremely high accuracy standards, performing emotional and relational labor, developing machine-like thought processes, and facing financial instability alongside strict organizational control. An emerging form of managerial control in AI production, termed *hybrid management*, manages workers by both traditional factory-like criteria and gamified platforms, but they receive the benefits of neither. Additionally, we introduce the terms *context-specific accuracy* and *unifying accuracy* to foster a more nuanced and context-sensitive understanding of AI training.

This work makes the following contributions. First, we build upon prior research by introducing the concept of *precision labor*, which is informed by feminist scholarship on work and invisible labor with emotional [31], relational [3, 47], and immaterial [40] dimensions, as well as care and repair work [36, 46]. Precision labor refers to **the hidden and excessive labor involved in erasing the messy, ambiguous, and uncertain aspects of technology production to present technology as objective, truthful, and high-quality, even when such pursuit can be excessive, arbitrary, and harmful to laborers.** Precision labor provides a new lens for understanding the disproportionate impact of excessive and hidden labor as well as emergent harms experienced by digital labor communities. Second, we contribute an empirical understanding of how accuracy and precision are perceived, established, and managed in practice. This extends previous work investigating dominant preferences and values in ML studies [5, 8, 28, 45, 70] that focus on accuracy and precision. Our empirical findings illuminate how these values are constructed and legitimized, and suggest alternative approaches to promote worker agency and well-being. Third, we build on prior work on standard procedures and layered controls that shape data production [57] by incorporating machines as an additional layer of control and source of epistemic authority. Finally, we provide empirical insights on AI data annotation centers in China and their hybrid management. This contributes to global research on the challenges and power dynamics of data production [10, 57, 73, 90, 93].

2 Background and Related Work

2.1 Data Work and AI Training as Collaborative Practices

Researchers in social computing and related fields have long been interested in interrogating data practices [22, 62, 64, 76], . This research has revealed how data are constructed and crafted [15, 37, 63, 67]. The routine data practices, workflows, negotiations, and power dynamics between different actors involved shape the process of data production and the produced datasets [57, 68]. For

[2]The term *AI trainer* stems from *National Occupational Skill Standard for Artificial Intelligence Trainers* in China [58].

[3]Among different subfields of AI, what has been prevailing in public narratives and across diverse sectors in recent years is machine learning (ML), a specific and data-driven statistical approach to AI. Specifically, the development of ML systems that involve big data and larger classification systems requires massive data to train models, which will identify patterns in data so that it can be further used to make predictions and inform decision-making. Currently, the development of such ML systems still requires extensive data to be annotated to be readable to machines. In this paper, we use the terms AI and ML interchangeably.

[4]One area is formulating and implementing relevant skill-based household registration (*jineng luohu* 技能落户) in the context of Chinese household registration system (*hukou* 户口制度), along with associated social benefits for qualified AI trainers.

example, Miceli et al. [57] situate different actors in the standardization procedures and hierarchical structure within business process outsourcing companies. They not only control data quality but also impose their own interpretations of data. Gray et al. [26] and Whiting et al. [97] center data worker communities and identify their collaborative activities to manage administrative tasks, share useful information about tasks and requesters, and support others to complete tasks. Shifting the attention from technical workflows to the interactions with clients, Kross and Guo [38] identify the labor required to establish trust, accommodate clients' needs and constraints, co-framing problems, bridge gaps between data scientists' and clients' expertise, and provide clients with emotional support. Posada [73] further shows how local networks of data workers in Venezuela protect workers against the economic and social risks of disembedded markets and insufficient regulation. These collaborative activities also require adapting to organizational structures [85] and are conditioned by inter-organizational dynamics [68, 79].

Moreover, the trends of developing and applying (semi-)automation and AI-assistance technologies and tools in AI training and data production actually introduce nonhuman actors to data production process [14, 92]. The introduction of new actors has given rise to increasing research on how human-machine collaboration affects data practices [1, 94]. For instance, Ashktorab et al. [1] highlighted the additional labor, such as rechecking data when the quality of AI assistance decreases. Wang et al. [94] reported data scientists held mixed attitudes about the usefulness of automation tools. They felt the tools could enhance their efficiency and help produce "good enough" work to meet clients' requests, but also doubted the quality of work that existing automated systems and tools could deliver.

AI training's labor and collaborative practices are often rendered invisible by "an industry-wide strategy to de-emphasize the human factor in AI production" [65], to the extent that depict such work as ghost work. This echoes an increasing body of literature on the hidden and often unrecognized and underpaid labor in data production [16, 29, 35, 54, 69, 77, 86]. The strategic invisibility of labor imposed by technology companies not only perpetuates precarious working conditions for workers, but also conceals human subjectivity and intervention in data production [65]. As Passi and Jackson [67] aptly stated: "It takes *work* to make data work."

This paper explores collaborative practices in the production of ML datasets and AI training, and contributes to the growing literature on global data production [42, 57, 73, 90, 93] by adding a case in an understudied context: China. In China, AI training heavily depends on a local workforce, unlike in other countries where data work is often outsourced to a global workforce or assigned by international clients. While the outsourcing of data work in China is organized in varying ways (e.g., through crowdsourcing platforms [100], BPOs [99], or a combination of the two), AI data annotation centers (such as our field sites) have emerged as a significant site for data production. AI data annotation centers may work with major tech companies or leading AI data platforms, receive support from local governments, and expand their operations into developing regions, including urban cities and rural villages. In this context, institutionalizing AI training as a specific profession not only acknowledges the surging demand for this growing workforce [5], but also seeks to standardize skills, build a professional workforce, and foster workers' career aspirations. We define a specific form of hidden and excessive labor, driven by clients' preoccupation with accuracy and precision, which we refer to as *precision labor*.

We argue that the pursuit of accuracy and precision in data production can be pushed to such an extreme that it leads to excessive labor and harms AI labor communities. Moreover, our research informs the ongoing discussion about human-machine collaboration in data production [1, 94] through the lens of data workers. While data scientists may perform part of data work [63], they are internal research team members and are trusted for their authority and credibility over the data practices. In contrast, AI trainers, located in the external sourcing process and at the lower end of the AI value chain, are blamed for introducing bias. They are subject to the judgments of actors with more epistemic authority, and their expertise remains underappreciated [56, 57]. Instead of a generative and collaborative human-machine relationship, this study shows that AI trainers are subordinated to machines, and that machine evaluation and output constitute an additional layer of control and source of authority in data production.

2.2 Accuracy and Precision in and Beyond AI Training

Accuracy and precision are crucial scientific values [39], among what Longino [49] refers to as epistemic values that "are taken to conduce to truth or rational belief". As Wise [98] puts it: "Precision is... responsible, nonemotional, objective, and scientific." However, accuracy and precision, as ubiquitous constructs, are difficult to pin down with unifying definitions. Historically, the notion that "precision implies accuracy" has been recognized and well established [98]. In line with this perspective and for the purposes of this paper, we propose that precision encompasses accuracy and, in specific contexts (such as AI training), the terms can be used interchangeably.

In the guise of being clear, objective, and unquestionable, accuracy and precision are often used to justify using and promoting technical products. For instance, accuracy is used to justify the US's growing investment in automated targeting systems. Suchman [83] points out that the accuracy of weapons striking targets can be improved since ambiguity persists in how weapons identify legitimate targets. What is a "legitimate target" is always constructed and shaped by politics. Similarly, in the context of AI deployment, high accuracy is used to justify and promote the use of AI products [28]. However, improved outcomes from accuracy metrics do not resolve the ambiguity and uncertainty in how problems are defined and targeted objects are constructed and measured, such as in emotion recognition [81] and risk assessment [27].

While accuracy and precision can be presented as unambiguous and incontestable facts, upon closer inspection, they are contextual and subject to human discretion, disciplinary conventions, cultural values, and the social and political shaping of technologies [51, 98]. Wise [98] also emphasizes the politics of establishing precision, as

[5] In 2024, AI trainers have been included in the list of occupations of short supply and high demand in several Chinese provinces (e.g. Hubei province) and cities (e.g. Guangzhou, Shanghai, Beijing). For example: 朝阳首发重点产业紧缺人才目录 https://rsj.beijing.gov.cn/xwsl/mtgz/202409/t20240906_3791215.html

the process of reaching a collective consensus on what is perceived as true and certain is shaped by power dynamics. Centering the production of scientific knowledge, Longino [48] further proposes the term "empirical adequacy" over accuracy to highlight context specificity, as well as the flexibility and subjectivity in interpreting accuracy, which is susceptible to socio-political influences and practical feasibility.

Accuracy, together with efficiency, universality, impartiality, and scale, is among the prevailing preferences and values in ML communities associated with "success" and "progress" [5, 78]. The accuracy performance metric is a salient evaluation criterion for the performance of models and AI products. It involves "performative, simplified, normative, and rhetorical numbers to convince others" [28]. Grill further demonstrates the process of producing and constructing accuracy and uncovers how ignorance is strategically produced to conceal the contingency and flexibility in the production of accuracy and the resulting disproportionate harms to marginalized communities.

Accuracy of datasets has long been regarded as an important dimension of data quality [2, 18, 96], yet does not have a widely recognized definition [18, 91]. Data quality requires a variety of accuracy metrics with different definitions and operationalizations [18], and is also specific to contexts and application scenarios [7, 66]. Moreover, accuracy in data production is usually defined and standardized by clients, requiring rigorous and excessive labor from workers to comply [57]. Clients monopolize the goal and measurement of accuracy, and workers "usually obey everything that [clients] say" in cases of uncertainty [57].

In line with previous literature that examines and questions the values in ML studies [5, 28, 45, 70], we examine how accuracy and precision shape data production and AI training practices. We contribute to existing work that deconstructs accuracy [28, 83] by describing the processes to construct and manage to ensure accuracy and precision. We also present the extensive labor necessitated to perform them in AI training. Echoing existing work on precarious working conditions in data production [57, 89], we show the disproportionate harms for workers to achieve such accuracy and precision standards. We also highlight how excessive labor becomes normalized and legitimized in the name of accuracy. We question whether the goal of high data quality can adequately justify the obsession with accuracy and precision in producing commercial ML data. We also call for weighing the pursuit of accuracy against the resulting harm to workers and fostering sustainability within the ML data sector.

3 Field Sites and Methods

Ethnography is particularly apt for studying work and labor because it highlights the "interpretative flexibility of technological artifacts" [72]. It allows us to understand the intricacies, significance, and politics surrounding AI training and data work. On a micro level, ethnography "systematically restores the value of individual experience, and takes seriously the notion that how individuals perceive the world" [82]. This project combines multi-sited in-person fieldwork [53] and virtual ethnography [84], as interlocutors and AI data annotation centers are distributed across different regions of China, and diverse digital systems are embedded in workers' everyday lives. To explore Chinese AI trainers' motivations, aspirations, and everyday professional experiences, and understand the effects of AI on their work experiences, we relied on multi-sited ethnographic fieldwork conducted in two business process outsourcing companies located in underdeveloped regions in China. The first BPO, which we pseudonymize as XiaochangA[6], is nestled in an AI data annotation center at a technology park in Central China. It specializes in producing ML datasets and provides tailored services that align with the strategic AI-related goals of DachangX[7], one of China's leading AI companies. The second BPO, XiaochangB, is located in a "maker space" in Southwestern China, and also focuses on ML dataset production but primarily works with one of China's most prominent data crowdsourcing platforms.

On the ground, the first author relied on grounded theory to guide their research [12]. In January 2023, he started in-person fieldwork in Southwestern China by asking broad questions about data work, AI training, and labor. Located in an area of Guizhou province known as China's "Big Data Valley" (*dashuju gu* 大数据谷), [8] this field site in Southwestern China provides an ideal location for fieldwork. Early in this exploration, accuracy and precision emerged as key themes, especially when talking with interlocutors such as managers and workers. They frequently mentioned how their training, managerial strategies, and goals were intrinsically tied to these values. Concurrently, the first author routinely met with the last two authors to discuss his observations and emergent themes. He then used theoretical sampling to gather more data related to this topic and expanded the fieldwork from XiaochangB to XiaochangA. The first author initially planned to expand fieldwork to another Southwestern Chinese city associated with a different major technology company. However, after access was denied, he redirected his research to the second field site in the technology park, where a robust data industry in Central China had flourished since 2019 [9]. He drew on his professional networks and cultural knowledge to secure access to research sites. The selection of field sites in Southwestern and Central China was guided by his research goals, theoretical interests, and practical considerations such as time, financial constraints, and field access.

During nine months of fieldwork in China, the first author actively participated in numerous activities to gain a deeper understanding of the lived experience of AI trainers working for the data industry in China. These activities included undergoing a three-week-long virtual AI training course and examinations, personally attending and observing major industry events such as the 2023 China International Big Data Industry Expo (Big Data Expo) [10], and observing media interviews and news productions related to

[6]The term "Xiaochang" （小厂） in Chinese refers to small factories. We have chosen "Xiaochang" as a pseudonym to honor the Chinese context in data work and AI training.
[7]"Dachang" （大厂） in Chinese denotes large factories and is often employed colloquially to describe major tech corporations in China. We have selected "Dachang" as a pseudonym to more accurately reflect the study context while ensuring the anonymity of the participants.
[8]Guizhou: The Big Data Vally of China: https://sponsorcontent.cnn.com/edition/2018/guizhou/china-big-data-valley/
[9]Faces for cookware: data collection industry flourishes as China pursues AI ambitions https://www.reuters.com/article/economy/faces-for-cookware-data-collection-industry-flourishes-as-china-pursues-ai-ambi-idUSKCN1TS3E7/
[10]It is claimed to be one of the world's first big data expo by the organizers. https://www.bigdata-expo.cn/?lang=en

China's data industry. To fully immerse themselves and gain a first-hand perspective of everyday work routines and practices, he worked as an intern for eight weeks at XiaochangA between March and June, 2023. In this role, he worked with other AI trainers from 9:00 AM to 6:30 PM, six days a week. His work involved hands-on participation, observing different data projects and tasks, attending team meetings, and joining colleagues for daily lunches at the technology park canteen. Outside of official hours, he fostered deeper connections with interlocutors by visiting their accommodations, meeting their family members, dining with them, and taking short trips to local tourist attractions with them. When speaking with interlocutors, he disclosed his background and position as a researcher from a US institution.

The first author's identity as a former journalist and Chinese citizen with a strong network in the country helped him gain field access. In addition to in-person observations, he gained access to internal documents such as instructions and requirements related to AI training and data work. He also obtained additional online data through engagement with online working group chats, screenshots of different work tasks, and virtual training/meeting transcripts. He further incorporated insights from 16 in-depth interviews with employees. Before each interview, He informed interlocutors about the study's goals and confidentiality measures and obtained consent for audio recording. At the beginning and end of each interview, he emphasized that interlocutors could request data deletion without penalty. The interviews explored topics including interlocutors' lives before and after becoming AI workers, their motivations, migration experiences, daily use of ICTs (including AI annotation systems and social media), experiences and challenges of working as AI workers, and their future plans. As accuracy and precision emerged as significant themes during the fieldwork, he incorporated interview questions about interlocutors' perceptions of these values and how accuracy and precision influenced their daily work experiences and well-being. Interviews were conducted in Chinese and then transcribed verbatim. The quotes used in this paper were translated into English by the first author.

Our data also includes over 190 pages of field notes. Drawing on Emerson et al. [21]'s approach, the first author used a three-column structure when generating field notes, enabling systemic data collection and reflexive analysis throughout fieldwork. Most of the field notes were written in Chinese to capture the subtleties and nuances of the language and interactions.

Our data is grounded in interviews, observation, field notes, and online archives. We used a grounded theory approach to analyze data [12]. The coding process was conducted in English, with in-vivo codes retained in Romanized Chinese (*Pinyin* 拼音) accompanied by English explanations. This approach preserved cultural nuances. We first used open coding to inductively identify themes within the data in ATLAS.ti., then progressed to axial coding, in which codes were organized around points of intersection [12]. Examples of codes include unpaid labor, punitive measures, alignment with clients' agendas, human-machine entanglement, thinking like AI, and accuracy rate as a management tool. We used a visual and collaborative digital tool, Miro, to consolidate, group, and visualize relationships between codes and themes, enabling collaborative refinement of our findings. We reached shared understandings through frequent discussions and weekly meetings during the process, which helped resolve analytical disagreements and generate the themes reported in the results section. Additionally, we anonymized all data in this paper and used pseudonyms for interlocutors, companies, and locations to protect interlocutors' privacy and ensure their safety. This study was approved by the first author's university's ethics review board.

By reflecting on how our positionalities and identities influenced our data collection, analysis, and writing processes, we emphasized the social contexts in which our research took place. In this section, we focus primarily on the first author's positionality (because he interacted directly with interlocutors and led data collection and analysis), but the authorship team includes researchers with a diverse range of genders and nationalities, including both Chinese natives and people from several other countries. All team members are highly educated, affiliated with Western universities, and specialize in social computing, with a strong emphasis on the social impacts of technology. The first author's family background, education, and migration experience positioned him as both an "insider" and "outsider" in this study, which shaped the direction of the work, data collection, and analysis. As a Chinese native who grew up and received early education in rural China, he shared important cultural and linguistic foundations with this study's interlocutors. Like many interlocutors, he migrated for work and experienced extended separation from his family. These shared experiences helped build rapport and trust between the first author and his interlocutors. His ability to relate to cultural references, social norms, and lived experiences in China facilitated deeper connections with this study's interlocutors. Additionally, his previous experience as a journalist at a renowned Chinese news outlet enhanced his sensitivity to institutional power dynamics in Chinese settings. However, his position as a highly educated researcher at a U.S. university, Western academic training, and years of living abroad made him an outsider among interlocutors. In the field, interlocutors often asked him about China-U.S. relations and his post-graduation plans, highlighting their outsider status. Throughout the research process, the first author actively reflected on how these positions improved his work through transparent communication with interlocutors about his background and research goals and maintained a separate research journal to document his feelings and potential biases [71]. The categories of "insider" and "outsider" remained fluid as the first author's personal, professional, ethnic, and diasporic identities evolved. In studying data-driven technologies, he focused on technology's material impact and human agency, believing that while technology can address certain social issues, it cannot resolve fundamental social problems. These views on technology, combined with his education and lived experience, shaped his research approach and knowledge production.

4 Findings

4.1 Perceiving and Establishing Accuracy

Our findings reveal that accuracy can be context-sensitive and carry multiple meanings. Workers internalize the importance of accuracy and precision without necessarily understanding why they are significant nor the mechanisms behind their influence on models and AI systems. They often accept extreme accuracy standards imposed by clients as given.

4.1.1 Perceiving Accuracy. Many workers at the AI data annotation center believed that the AI training work they do refines Al systems, yet they understand little about the mechanisms behind AI. Notably, they were often told that their work should be done accurately and quickly. Yiran was a stay-at-home mother who wanted to work while taking care of her children. She first encountered this profession through a relative in her village near the technology park. The 33-year-old working mother recalled she had no idea what data annotation and AI training were:

> *I was told that it's roughly about annotating data for AI to learn because when people create a machine, it doesn't know anything; it needs to rely on absorbing data.... For example, if you want machines to recognize people, how can they know it is a person? You will need annotators to correctly mark this person out correctly and identify features of this person, such as head.... This also applies to other contexts.*

Narratives about doing things correctly and accurately were often emphasized even before the employees started working as AI trainers. What Yiran was told about creating accurate data for machines to learn was similar to the experiences of many other workers. They held a widespread belief that accuracy is vital for this type of work. When asked about why workers need to perform certain specific types of tasks, such as correctly identifying traffic lights for a project related to autonomous vehicles, Haoyang, a 22-year-old man, explained:

> *[It is related] to AI. Just like Meituan [a Chinese shopping platform] has these robots for food delivery, they need to recognize things such as traffic lights. If they recognize red lights, they will stop. Machines need our help as we implant our knowledge in them so that they can make correct judgments. We are the transmitters. We transmit our knowledge about the traffic rules to them so they can accurately recognize things.*

4.1.2 Establishing Accuracy. In AI training, the term "accuracy" can have multiple meanings, often depending on the specific context, as illustrated in Figure 3. For instance, different tasks and projects involving voice recognition, computer vision, and generative AI technologies could have distinct measurements and requirements. Nevertheless, across projects, there was a common understanding of accuracy based on the result of standardized statistical measure, as shown in Figure 3. This common understanding, which we call "unifying accuracy", refers to the overall accuracy rate of projects. Workers were often arbitrarily required to achieve an accuracy rate over 95% and sometimes reaching as high as 98%. For example, in a computer vision project for autonomous vehicles, Haoyang's labor was measured by the extent to which he correctly categorized the instructed objects. The images captured during the day have much higher quality and visibility than the evening ones due to lighting conditions and limited camera technology, such as low sensor sensitivity, but the standard did not change based on the time that images were taken. As a result, he performed additional labor, such as brightening the images and going through each image multiple times, to achieve the expected level of accuracy.

Moreover, Haoyang and his colleagues were instructed to annotate those objects with 2D bounding boxes, which needed to be strictly fitted to the objects with less than a 3-pixel fitting error. He was told repeatedly that he should pay specific attention to this requirement, which they call *Tiehe* (贴合), one of the key elements to measure the accuracy of these tasks. When Haoyang and his colleagues first started, they magnified these images multiple times to draw precise bounding boxes around objects like traffic lights. They were told that the accuracy would be compromised if the Tiehe was not precise. Haoyang was one of the most prolific AI trainers on the team. After a few weeks, he believed he could ensure Tiehe without magnifying images. This meant he could spend less time accurately drawing bounding boxes, which directly translated into improved productivity. "Since I've done so many [tasks], I can simply draw a box to fit the object perfectly. Some of my colleagues still need to magnify the image several times before drawing polygons and then adjusting it back and forth until it fits precisely," he added. Even though Haoyang could rely on his tacit knowledge to draw bounding boxes more efficiently, he was still instructed to complete them in standardized ways, such as magnifying images multiple times, causing the tasks to be even more mechanical and tedious. Such labor is often unrecognized and excessive.

Despite the multiple meanings and the contextual nature of accuracy, for many workers, accuracy was more of a homogeneous and arbitrary production goal, manifested through the clients from technology companies setting demands for unifying accuracy and reinforced by different actors. Zimo, who has worked as both an AI trainer and reviewer, revealed: "Accuracy means meeting clients' requirements. some are set at 98%, others at 99%. The client will approve your project if you achieve their accuracy expectations." He added that he has even heard of some "high-end" better-paying projects where clients specifically ask for a 100% unifying accuracy rate.

Our findings, grounded in the Chinese commercial context, reveal that clients had consistently high accuracy expectations of a typical minimum threshold of 95%. Some "high-end" (better-paying) projects demanded even higher standards, reaching 98–99% or occasionally even 100% unifying accuracy. Although the variation in accuracy requirements was not substantial, the labor implications of these differences were significant. Through observations and conversations, we learned projects demanding extreme accuracy levels often made additional demands. Their demands included more stringent context-specific accuracy requirements, additional rounds of revisions, increased frequency of quality checks, and extended unpaid waiting periods for feedback. Despite promising higher payment, such projects were more likely to be *Huishou* (回收), where workers are likely to end up with less pay. In the context of AI data annotation centers, *Huishou* involved rejecting AI trainers' work at different levels: rejecting specific tasks, excluding workers from an entire project, and rejecting all tasks in project segments. As the level increased, the financial losses for workers became more severe. This further illustrates that clients obsessed over performative accuracy without carefully considering the feasibility, financial costs, and potential labor and economic harms, such as wage theft behind such requests, which we will present in Section 4.2 and elaborate on in Section 5.

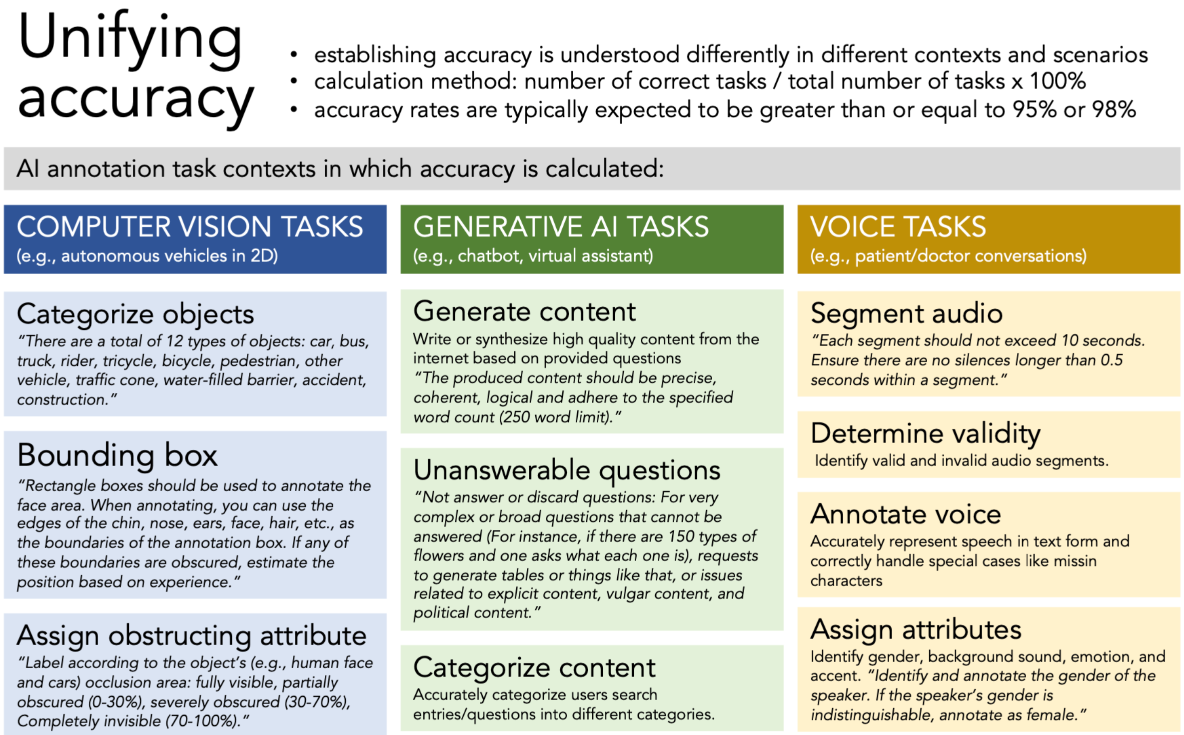

Figure 1: The establishment of accuracy is understood in different scenarios and dimensions. This figure synthesizes the instruction manuals collected by the first author, with selected quotes serving as examples. Texts with quotation marks are a direct translation from the instruction manuals.

4.2 Managing and Making Performative Accuracy

Previously, we documented how technical accuracy was perceived and established. This section highlights how accuracy was managed and intertwined with power dynamics. Our findings illuminate that an excessive focus on high technical accuracy can be labor-intensive and cause multiple harms to workers. Specifically, workers engaged in excessive training and revisions, performed invisible labor, and developed machine-like thought processes, all while facing financial instability and strict organizational control.

4.2.1 Excessive training. To meet clients' specified accuracy standards, workers often were trained in excessive ways. Extensive training for data work was a common practice observed in both field sites. In most cases, workers were trained over an initial phase lasting from three nine-hour days to a few weeks before working on the projects. Their training typically involved studying general training materials, participating in video conferences, completing designated practice assignments, and undertaking assessments covering foundational concepts and theories and associated techniques. This training was often mandatory and unpaid. Yichen shared his training experience at DachangX:

> *When DachangX trained us, we began with 2D bounding box annotation. It was for autonomous driving recognition, requiring us to accurately identify all road obstacles. We marked everything, from pedestrians and bicycles to cars and crash barrels. Each object had specific attributes that had to be labeled correctly. For instance, we'd differentiate between walking pedestrians and those lying down. If there were a chair, we'd note whether someone was sitting or lying on it. Notably, babies on strollers couldn't be labeled as pedestrians.*

Before taking on a new project, workers often went through unpaid new training sessions due to different rules and specifications. This subsequent training was frequently shorter and acquainted them with project-specific guidelines. This often entailed participating in video or voice conferences facilitated through platforms like VooV Meeting, where either clients or managers clarified the regulations. Notably, these meetings were often impromptu and scheduled at the last minute.

According to the interlocutors, their training could span a week. However, during the first authors' fieldwork, the training duration was shortened to 1–2 days. One reason for shorter training was the absence of incentives for participation. Since workers possessed foundational knowledge from their initial training and previous work experience, the subsequent training often contained redundant content. Some workers considered further training rounds a waste of time, especially since they were unpaid. Consequently, as the first author observed, workers sometimes became inattentive during video sessions, diverting their focus to non-work-related activities such as chatting with colleagues or browsing short videos on social media.

In certain regions, the first author observed that managers required prospective workers to present an AI training certificate before hiring them in the name of enhancing data quality and job security. Obtaining an AI training certificate required undergoing weeks of formal training, taking a strictly scrutinized exam, and spending over a month of salary on fees, often paid by individual workers themselves. Like many AI trainers, the first author personally invested several weeks and around $440 (USD) to go through this formal training process. After talking with managers, he later found the certificate was "unnecessary" for doing such work. However, managers could present the annotation teams as skilled and professional, and promise they would provide high-accuracy data, which served the clients' interests more than the employees'.

Taken together, while excessive unpaid training may be perceived as an essential step for achieving optimal performance accuracy and high data quality, it often proved excessive, harmful, and counterproductive for workers in practice.

4.2.2 Homogeneous workflow. To ensure high accuracy and data quality, a standard workflow was often established across different projects, involving stakeholders like clients, intermediaries, and service providers. The clients relayed data to intermediaries, who then delegated tasks to the service providers. Within the service provider's realm were AI trainers, reviewers, quality inspectors, and managers. For instance, the AI company DachangX might have acted as the intermediary (project management), while the service provider could have been a BPO.

Figure 2, adapted from a company figure and used widely in the industry, shows a flowchart designed to ensure perceived high accuracy. Once AI trainers completed their data annotations, they submitted their work to reviewers. If they identified multiple errors or believed the accuracy was below the threshold, they returned the data to the AI trainers for revision. Once the task passed the reviewers' standards, quality inspectors conducted additional accuracy checks. With successful clearance from both reviewers and quality inspectors, the annotated data could be submitted to the client's project overseers. This team assessed data accuracy and sometimes employed algorithms to test the data, ensuring that the model performance met high accuracy expectations.

If any stage of this verification process was unsuccessful and multiple revisions were requested, the data was either returned to the service provider, which resulted in pay deductions for workers. If the data work was deemed a failure, it was called *Huishou*. When AI trainers were rejected through *Huishou*, the rejected tasks were often redistributed as new tasks/projects, often to different teams or companies. While these recycled tasks were typically priced lower than brand-new ones, workers tended to view them as more manageable. Specifically, interlocutors believed such tasks were more mechanical and intuitive, so they could complete them faster with lower effort, potentially earning more money through a higher volume of completed tasks. This practice reveals that the client company was cutting operational costs by creating excessive work. They devalued human labor not only through direct task rejections in the name of accuracy but also by redistributing rejected tasks. In cases of specific task rejection, workers were uncompensated for their time and effort. Haoyang was furious after 6 out of 8 tasks led to *Huishou*. The tasks were valued at approximately 60 yuan ($8.35) each and he worked very hard to complete them in a timely manner. In theory, he should have received 480 yuan, but in reality, he only got paid for two tasks, amounting to approximately 120 yuan. He remarked:

> *I was so frustrated. I even contemplated resigning, you know?... Even if I made multiple mistakes, they could have pointed them out individually. I would accept and learn from them. You can't just reject almost all of my work at once.*

Moreover, iterations of the revision often resulted in workers being removed from projects and losing all payment even for the approved tasks. Haoyang elaborated: "If your tasks are repeatedly returned for revision, you could end up losing all of your earnings for that project." The highest level of *Huishou* observed by the first author occurred in large projects, where a segment of the project was deemed a failure, causing all workers in that segment to lose their payment. Sometimes, this was beyond the control of individual workers and was interpreted by the first author as a form of collective punishment. The risks of *Huishou* of project segments created additional pressure for workers to engage in extensive labor to ensure accuracy, as their performance affected not only individual payments but potentially the earnings of all workers in the segment.

Despite the risks of *Huishou*, many interlocutors continue revising their work in the hopes of maximizing their earnings. This process was often emotionally taxing, especially when workers realized that hours or even days of their labor were ultimately wasted and went unrewarded. "Mentally, it's draining," Yiran added when undertaking tasks with higher standards. "It feels like every day here is just a futile hustle."

Zehong explained the dilemma of deciding whether or not to revise previous tasks to improve accuracy: "Sometimes we have to revise [tasks] up to four or five times. Beyond that, the task is considered a failure.... [But] if you don't revise them, they also get marked for *Huishou*. You won't receive any payment... It becomes a vicious cycle." Later on, he discovered that this standardized workflow and the emphasis on accuracy provided justification for intermediaries and clients to cut costs by underpaying workers. In other words, they sometimes set exceptionally high standards, knowing workers would fail repeatedly. He added that many of them would get very angry and even use curse words while talking with other colleagues in private. Gradually, for projects that demanded very high accuracy, he started limiting his revision efforts and instead waited to be switched to new projects. "Even if I do it, there's no guarantee of money, so why bother keep doing it?" If all goes well, the project concludes, and the workers will finally be compensated.

4.2.3 Power asymmetry in establishing accuracy. The first author also regularly observed power imbalances that manifested at different stages of the AI training work. These imbalances contributed, in part, to hierarchical and top-down decision-making and emphasized performative accuracy. Apart from the unclear and constantly

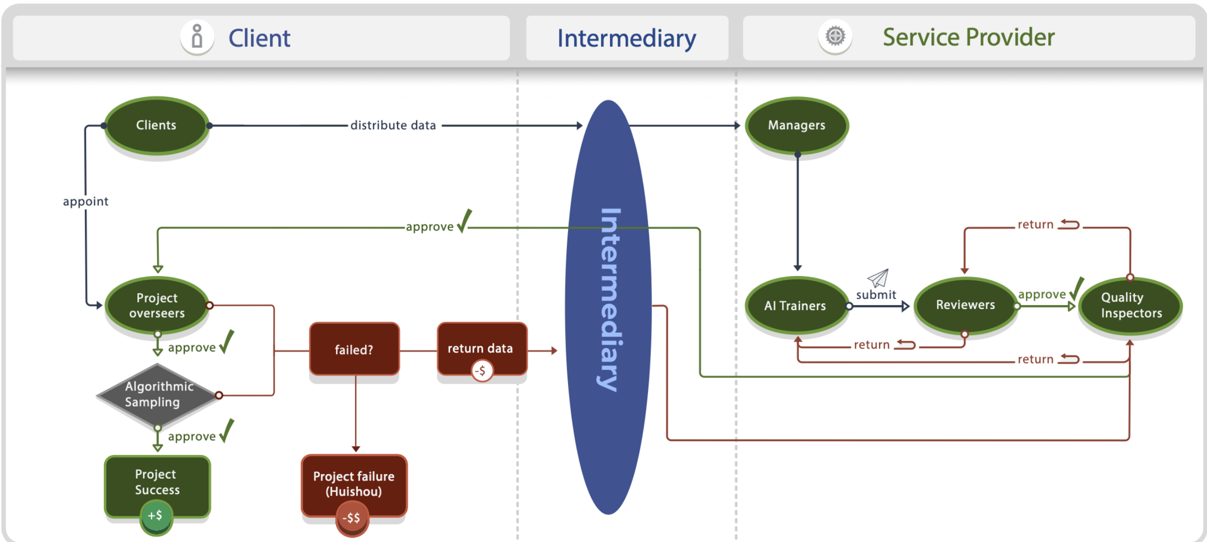

Figure 2: An emblematic example of workflow in AI training projects to manage and aim at ensuring high accuracy. The figure is adapted from a company diagram and is widely used across the industry.

changing task instructions, power imbalances permeated the interactions and dynamics among all human actors involved: clients, intermediaries, AI trainers, reviewers, and quality checkers.

The first author observed many different projects and several scenarios at the technology park. It became evident that while power dynamics existed among the AI trainers, reviewers, and quality checkers, the ultimate authority resided with the client. This observation echoes findings presented by Miceli et al. [57]. Specifically, clients determine the high accuracy rate, establish annotation rules, and have the power to change and interpret the specifications of the rules. Within this hierarchical structure, AI trainers often felt compelled to follow the guidance of reviewers. Similarly, reviewers felt obliged to respect the views of quality checkers while also adhering to the rules set by the clients. Zimo, a reviewer, revealed that despite the perceived authority and tacit knowledge he had in the eyes of the AI trainers, he felt his influence was very restricted. The rules created by clients limited his decision-making regarding what was accurate. He described the nuances of these power dynamics of this process:

> Researcher: *When you serve as a reviewer, do people take your words as rules?*
> Zimo: *No, reviewers also follow the rules.*
> Researcher: *If, as an AI trainer, I consult you about the acceptability of certain annotation approaches and you approve, I'd assume they're acceptable.*
> Zimo: *That's not true; as reviewers, we don't establish the rules; we reinforce them.*
> Researcher: *OK, I see.*
> Zimo: *When consulted, we assess based on the rules, if a particular annotation style is acceptable or it is the correct way to do it.*
> Researcher: *The problem is that the rules we see aren't always concrete, and they change.*
> Zimo: *If it's ambiguous and uncertain, you can escalate the issue. Above us, there is the client's project overseer (xiangmu laoshi). Whatever they say, we'll follow their guidance.*
> Researcher: *So, does that mean the project overseer's directives could be viewed as rules?*
> Zimo: *Yes, that's correct.*

Perspectives like Zimo's regarding rules and authority are not isolated; they were a norm in the technology parks. While reviewers and quality checkers could only make decisions based on already established rules, they still had considerable perceived authority over the AI trainers. This was primarily because AI trainers frequently interacted directly with reviewers and relied on their feedback. Within the service provider, there was a palpable power imbalance between AI trainers and reviewers and between reviewers and quality checkers.

Indeed, when faced with disagreement among AI trainers, reviewers, and quality inspectors, the "ground truth" often came from the clients. Zimo elaborated:

> *Even if you're a reviewer, you still need to defer to the quality inspectors. However, at the end of the day, the clients' words matter. Whatever they ask for, you implement those changes; that's the only way to get your tasks approved.*

This means AI trainers often found themselves getting feedback from different sources and having to agree with almost all of them, leading to multiple rounds of revisions. Notably, agreement was more of a form of conformity in performance among workers at the lower levels of the data production hierarchy, namely, AI trainers. Workers consistently pointed out that the client dictated the desired accuracy rates and the rules, even though they might seem arbitrary or contradictory.

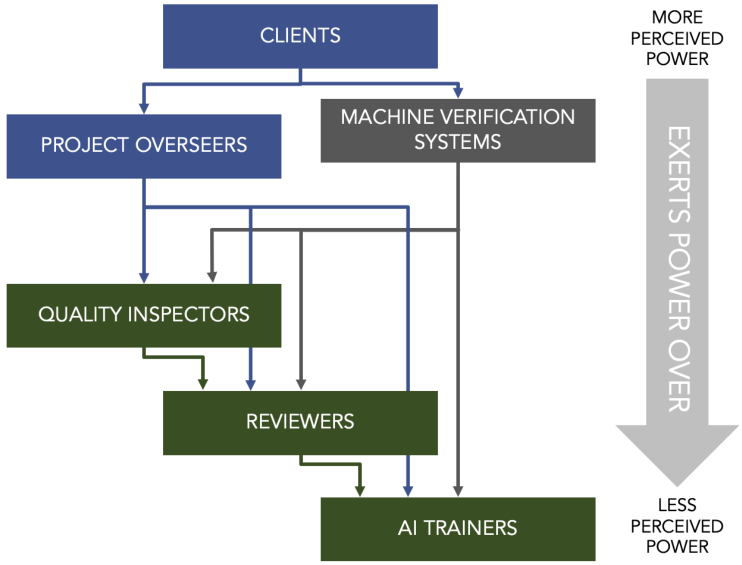

Figure 3: This figure maps the different power dynamics among different actors, including clients, project overseers, machine verification systems, and related workers.

4.2.4 Hybrid Management. We refer to hybrid management as a hybrid form of labor control and management, which is an apt characterization of our study's context. Workers were controlled in a setting with attributes typical of both factories and gamified platforms. Specifically, this study presents a case of hybrid forms of control, as workers were controlled and managed by traditional and factory-like criteria and gamification but did not receive the benefits of either. This was especially true regarding the income security and social benefits in traditional settings and the flexibility in the platform settings.

Similar to the factory setting, DachangX implemented a facial recognition clock-in system and mandated workers to use it to record their attendance. The first author observed workers routinely lining up in front of the computers equipped with the clock-in system around 9:00 AM. and 6:30 PM. to register their arrival and departure times. The record both captured attendance, which qualified workers for the 200 yuan ($27.80) monthly "perfect attendance award". Such incentives are important, as BPOs often have a strong desire to retain workers to ensure consistency in task performance and data quality. Additionally, cameras were installed in most corners of the building. While some workers accepted this surveillance system as a part of the office protocol, others deemed it intrusive. Importantly, these facial recognition-empowered clock-in systems have made skipping or being late for work much more challenging.

Yichen was an AI trainer who recently transitioned from working as a factory worker at an electronics factory in eastern China. During our interview, he expressed his dissatisfaction with the facial recognition-enabled check-in system:

> *I'm quite resistant to it because I just left a company with a clock-in system, and now I'm being asked to clock in again. I feel restricted. I really don't like it, but there's nothing I can do. We have to adapt to the environment as the environment can't adapt to us. That's how it is.*

Such clock-in systems ensure that workers physically work in the AI data annotation center for a prolonged period, even though their pay is often based on piecework. Workers tolerated spending idle time to get the monthly bonus and be assigned to new tasks. This idle, unpaid time could amount to workers spending days or even weeks in the center not actively working. In addition to waiting for new tasks, workers also waited for feedback from other actors and for rules to be established or confirmed. To achieve performative accuracy, some clients assumed workers must spend a certain amount of hours on given projects. Even if they finished early, workers still needed to ensure that they logged enough screen hours, resulting in additional unpaid waiting. These managerial practices, purportedly in the name of accuracy, directly increased waiting times and financial precarity for workers.

When joining a new project, workers often watched synchronous or asynchronous tutorials as training, which they perceived as crucial for improving data quality and accuracy. They were also invited to join online project groups where reviewers and quality inspectors highlighted workers' errors. Some workers were uncomfortable with their mistakes being highlighted in such a public manner. Yiran further elaborated:

> *If a reviewer thinks you made some annotation errors, they'll mention it in the group chat, often suggesting, 'If you're unsure, consult colleagues or check the rules.' It feels a bit like being criticized, like how a teacher would reprimand you in front of a class during school years.*

To enhance workers' performance and motivation, their performance metrics were mostly made public among their peers. Almost every day in the group chats, workers were provided with an Excel spreadsheet exported from the annotation systems detailing everyone's performance metrics. Performance spreadsheets included specific information such as username, project name, number of annotations, accuracy rate (often with approval and failure rates from both reviewers and quality inspectors), and hours spent annotating.

One late afternoon in mid-May, following the spreadsheet's release, activity surged in one of the WeChat groups. Manager Weinan commended a male colleague using the word "Niubi," meaning impressive. He was 50% more productive than his teammates and by far surpassed Meilian, a productive female worker known for consistently producing high-quality work. Strategies like public praise motivated workers to enhance their output or at least maintain above-average performance. Yiran conveyed her sentiments: "When you see others achieving so much, even if you doubt your capacity to match their pace, you're very likely to push yourself harder. You feel you can't let yourself fall behind; your performance needs to be at least acceptable, without lagging too far behind." The first author felt the same way; upon beginning a new project or facing a day of low productivity, seeing the ranking of the spreadsheet made workers feel like they were in the spotlight. Like Yiran, he would redouble their efforts the next day, which often meant sitting in front of the computer for even longer hours.

The competition among the most prolific workers were intense. The day following the manager's announcement, Meilian remained troubled. She regretted having taken an extended nap the previous day at her work desk, which she felt diminished her productivity. Over lunch in the technology park canteen, Meilian confided in the first author her ambition to be the first in the group — a sentiment she hesitated to express in the presence of other colleagues. Consider this exchange the first author had with Meilian about the impact of gamification and the dynamics of competing with male peers:

> *Researcher: Why are you competing with him?*
> *Meilian: He has the highest number.*
> *Researcher: Why do you want to be at the top?*
> *Meilian: Why can't I be?*
> *Researcher: Of course you can! You're very competent.*
> *Meilian: Do you know? He came in early in the morning. I started working on the tasks at 9:30 AM., but he started an hour earlier. At lunchtime, he orders takeout and eats them right away [in front of the computer].*
> *Researcher: Doesn't he have to handle things at home? Being out from early to late like that.*
> *Meilian: He's a man. What's there for him to take care of? His recorded working hours here are nine hours, while mine are just seven hours...I need to annotate quickly, annotate quickly.*

At the office entrance, public posters were hung on the door highlighting the achievements of these prolific workers. Unsurprisingly, the majority of them were men. Meilian found herself comparing her productively to others, especially to prolific male colleagues, even though many female workers like her were burdened with more domestic work at home, given women's expected role in Chinese households. Even brief diversions, such as restroom breaks, made her feel guilty, particularly when she observed her male colleague consistently seated at his desk. In response to these pressures, she started skipping her usual desk nap after lunch and resuming her work immediately.

While quantification and gamification techniques fostered productivity and uphold accuracy standards, punitive measures reinforced the company's emphasis on accuracy. These measures included pay deductions, suspensions, and mandatory unpaid retraining. While working on a computer vision annotation project, Yichen expressed dissatisfaction with the payment structure, which was heavily influenced by the accuracy rate: "For this project, if your task gets returned more than once, they deduct 20% from the commission. For every additional return after that, they would deduct 10%. That's why I didn't want to continue this project later." When asked how he felt about his work being constantly returned, he said: "I thought someone was targeting me and they were seeking revenge. Because each of us undergoes random quality checks, some people might feel upset when they see their data tasks being returned. As a result, they might return mine as well as a way of retaliation." Complaints regarding workers' tasks being repeatedly returned were common in the first authors' interactions with colleagues.

Every afternoon, Yuxi felt anxious before her previous day's accuracy rate was released. She explained that, similar to her colleagues, she was expected to maintain a high accuracy rate of 95% or above. If the accuracy rate fell below this threshold for two consecutive days, she could face suspension, meaning she would be barred from working on the project and would receive no payment. Persistent failures to meet the accuracy rate would require her to undergo unpaid retraining to remain on the project. Such management practices could be unnecessary, harmful, and counterproductive, which we elaborate on in Section 5.

After over two months of observations and talking with managers and colleagues, the first author realized that despite being excessive, managerial practices — such as the constant revision and punitive measures described above — were being imposed purposefully. They not only ensured performative accuracy but also could be used as a way to decrease labor costs.

Many labor-intensive and emotionally taxing efforts to enhance performative accuracy — such as excessive training, comparing accuracy with peers, and repeated task revisions — often went uncompensated, particularly among workers who were compensated on a piecework basis.

4.2.5 Machine subordination and humans as machines. To further improve the accuracy of machine learning models, algorithms were also added to the workflow (Figure 2). Machine verification systems are typically deployed by clients as an additional evaluation process, and the clients often maintain final authority over controversial results. Workers shared that the algorithmic evaluation often carried considerable weight in determining accuracy. When human actors such as quality checkers and project overseers were uncertain, "they would rely on machines to test and ensure the desirable outcome," Yuxi said. "It was really useful. What the machine says is what it is. We tend to go along with whatever it suggests." When the first author questioned that this approach seemed counterintuitive, suggesting that decisions about accuracy should be human-driven rather than machine-driven, there was a prolonged silence in response. According to Yuxi, there were cases in which workers appealed machine decisions. To make a strong case, workers needed to sit in front of the computers, investing time and additional labor to collect evidence. They researched official sources online, documented similar or related prior tasks with screenshots, and consulted more experienced colleagues. They also needed to wait for staff members on the client's side to look into the cases. In other words, the appeal process could be lengthy and yield few material benefits. Additionally, they often hesitated to challenge decisions due to existing power dynamics. This situation demonstrates how machine verification systems function not merely as technical tools but as instruments of power that reinforce hierarchies, amplify client authority, and compel workers to perform additional and often futile labor.

In addition, the pursuit of performative accuracy could lead to workers becoming subordinate to machines. While striving for optimal accuracy, workers' decision-making often becomes influenced by, and sometimes even subservient to, machine outputs, which are regarded as potential arbiters of the ground truth. To avoid being suspended and ensure a more stable income, workers frequently documented questions they answered incorrectly, individually or as a group. This documentation helped them understand and align with the logic that leads to performative accuracy. For instance, to understand the hidden logic of specific AI systems, Yuxi and her colleagues would test these systems and seek verification from the machine itself. Specifically, when moderating search engine content, Yuxi would use her phone to engage with the virtual assistant from that search engine, aiming to enhance its perceived accuracy and smartness. Though this virtual assistant is free to use, Yuxi had to download it to her personal smartphone and register as a user to test and verify responses. While she did not pay to use it, it still created additional labor outside her formal workflow. She had to use her personal device for work to use this virtual assistant over other mainstream AI tools and invest time in installation and testing. Such excessive labor is unrecognized and uncompensated in the formal process, but is vital to achieve high accuracy rates. While observing Yuxi in the crowded working space, the first author pointed out that her thought process seemed to mirror that of the machine, namely AI, rather than a human. She responded by mentioning she had been consciously training herself to think like an AI to ensure high task accuracy. The experiences and practices of Yuxi and her colleagues underscore that the reckless pursuit of accuracy could inadvertently encourage humans to emulate machine-like thinking and behavior.

Throughout the fieldwork, the first author rarely observed interlocutors negotiating with or resisting machine evaluations. Instead, he found them developing subtle ways to cope. Some created private chat groups where they could collectively document and make sense of their experiences. In these groups, they could work together to make their outputs more likely to be accepted by the systems. Some of them, like Meilian and Shiyun, got together to ask for transfers to a different project, but their requests were denied (management seldom approved such requests). As a result, most interlocutors simply tried to rush through and make do with their assigned tasks, hoping to move on to better assignments sooner. However, these strategies ended up reinforcing the machines' epistemic authority rather than meaningfully resisting or challenging it.

Structural barriers prevented interlocutors from resistance. Beyond the power imbalances in the workplace and punitive measures for missing accuracy requirements, as we mentioned earlier, workers' financial insecurity was exacerbated by weak social safety nets in developing regions of China. Many workers lived paycheck to paycheck, taking second jobs (such as food delivery and e-commerce package sorting) after their regular shifts. Such duties increased workers' likelihood of subordinating themselves to machines.

5 Discussion: Precision labor and its consequences

Our findings illustrate how technical accuracy is perceived, established, managed, and legitimized in practice in the context of AI training. In the following section, we connect interlocutors' everyday labor practices to the emerging harms driven by technology companies' relentless pursuit of technical accuracy. We contribute to the understanding of performative accuracy and critically examine its legitimacy. Additionally, we introduce the concept of *precision labor* to unpack the hidden and excessive labor involved in constructing and performing technical accuracy. Finally, we propose alternative approaches to enhance worker agency and well-being.

5.1 The Making of Performative Accuracy

Our findings document the relentless pursuit of technical accuracy by clients from technology companies. More importantly, they also illuminate how such a pursuit is materialized, legitimized, naturalized, and reinforced, perpetuating accuracy as a persistent value. We refer to this phenomenon as *performative accuracy* to highlight accuracy as a value that is constructed rather than self-evident and unambiguous. Specifically, we argue that these interpretations of accuracy not only constitute and establish what is considered accurate but also enact and shape accuracy within social realities.

Despite the aura of accuracy [51, 98], technical accuracy in AI training is context-specific, as we have demonstrated in our findings. Importantly, its standards and measures were **constructed through a set of discretionary decisions**. The degree of accuracy deemed appropriate and sufficient remained unclear, inconsistent, and unrealistic, leading to arbitrary and unjustified standards set by clients in practice, which often harmed workers. Additionally, the multiplicity of definitions and measures of data quality led to varying accuracy standards, further emphasizing the arbitrariness of their construction [18, 66].

We argue that **the making of accuracy is performative**. Performativity can be understood as iterative processes of producing specific knowledge and, simultaneously, shaping and enacting the reality that is represented. In turn, that reality reproduces and legitimizes this knowledge [6]. In the context of AI training, specific understandings of accuracy are formed when clients define what is considered accurate and set accuracy standards. They are further enacted through imposing and enforcing standards, which then require extensive labor from workers to follow and establish layers of control to ensure accuracy. Through these processes, arbitrary understandings of accuracy are naturalized and consolidated, fundamentally shaping the AI models produced downstream. Such uses could include but are not limited to autonomous driving systems, AI chatbots, and medical systems for clinical diagnostics; users of all of these may be negatively impacted by AI models trained on datasets that are supposedly precise but, in fact, hampered by the issues inherent in performative accuracy.

Performative accuracy serves to **naturalize the epistemic authority of actors** who define and shape it, **legitimizing the accuracy goals and standards** they set. Reaffirming findings from Miceli et al. [57], we found that in the Chinese context, clients also have the most epistemic authority and power to define and shape accuracy standards. Workers are inclined to perceive clients' interpretations of data as correct and self-evident. Workers are economically constrained from challenging their epistemic authority, as failing to meet the imposed accuracy standards could result in loss of income and employment. Importantly, our findings add to previous work on the epistemic authority in data production [57], by **highlighting the role of machines** in the accuracy evaluation process as **an additional layer of standardization and a new source of epistemic authority**. The integration of machines can further increase AI trainers' workload, for instance, by compelling them to maintain records of machine judgments as model answers to avoid punitive measures. Furthermore, compared to a collaborative relationship between data scientists and machines [94], AI trainers often find themselves subordinated to machine judgments, attempting to mimic machine outputs as much as possible to ensure high accuracy rates. Machine evaluation symbolizes accuracy and precision to such an extent that even in scenarios where machines and algorithms are solely used to identify invalid or falsified data, workers could still regard machine output as the arbiter of accuracy. This subordination of humans to machines not only highlights the epistemic authority of machines over workers, but also underscores the devaluation of workers' expertise. To ensure a desired accuracy rate, workers' judgments are considered trivial, their tacit knowledge is deemed unimportant, and their thought processes are expected to mimic those of machines.

In the production of ML data, achieving high accuracy is both a desirable and elusive goal. Examining the making of performative accuracy thus challenges the naturalization, legitimization, and enactment of specific understandings of technical accuracy. The process of producing and ensuring accuracy reveals power asymmetries imposed by arbitrary accuracy standards, highlights disproportionate harms faced by workers, and raises the question of whether such pursuits are more performative than substantive.

5.2 Implications for AI Production

Scholars have devoted considerable analytical attention to the harms stemming from humans' role in technology production, using terms such as *humans in the loop* [25] and *heteromation* [19]. In striving to fulfill the promise of automation, workers often engage in "ghost work" [25], in which their labor and contributions are purposefully rendered invisible [88]. Focusing on the invisible labor required to achieve extremely high accuracy standards, our findings highlight that AI trainers engage in different forms of invisible labor, such as emotional labor [31], relational labor [3], and care and repair work [46]. Specifically, to meet the high unifying accuracy rate, AI trainers must perform excessive hidden and unpaid labor, adapt their thought processes to be machine-like, internalize the significance of technical precision, and downplay the frustrations they encounter in the iterative process of reworking tasks. Moreover, AI trainers are subject to a hybrid form of labor management, which significantly intensifies workplace surveillance and control yet fails to provide the social benefits and protections often found in factory jobs, nor the flexibility typically associated with platform-based work. In addition, workers face wage theft, job instability, and financial precarity. They often need to invest excessive time and effort to engage in precision labor to secure their wages, as failing to meet accuracy standards carries significant financial consequences. We build on previous work on labor conditions and harms in AI production [35, 61] and different types of invisible labor, and extend it by conceptualizing the performativity of accuracy and discussing its two labor implications: performative accuracy can induce excessive labor for workers and further normalize and legitimize the labor and harms it induces.

Our findings show that the excessive drive for technical accuracy, though perceived by technology companies as essential and objective, not only results in hidden labor and harms for workers, but **such excessive labor can also be unnecessary for the quality of AI models**. Exploring the performative aspect of accuracy reveals the excessive labor demanded of AI trainers to meet extremely high accuracy standards, which may have minimal noticeable impact on the performance of AI models. This can manifest in excessive and rigorous standardization measures and in human subordination to machines, as discussed in Section 4.2.5.

Adding to existing work on extensive standardization measures in workflow and organizational contexts to meet accuracy standards [56, 57], our findings offer two additional insights. First, despite the availability of training, data companies can request workers to participate in certificate programs and obtain AI training certificates as a prerequisite for tasks. However, companies often do not compensate workers for the time spent in training and certification exams nor reimburse the fees workers incur. Additionally, the hybrid form of labor management fosters stricter and more intensive control than common outsourcing modalities in the data production sector, such as BPOs [56, 57] and crowdsourcing platforms [26, 73]. Hybrid management compels workers to engage in intense precision labor to secure payment, which in many cases justifies wage theft. In the name of accuracy, these two measures serve the interests of other actors at the expense of workers. For instance, through training certifications, managers can demonstrate the professionalism of workers and win clients over potential competitors. Clients benefit from such requirements and often use workers' failure to meet relevant standards as a justification to underpay them.

Moreover, meeting accuracy standards often requires subordinating human labor and judgment to machine outputs. Prior research on human subordination to machines [55] has focused on the harms and dehumanization of workers. We further argue that in the context of AI training, such subordination induces excessive labor, which is not only harmful to workers but also counterproductive for technology companies because it negatively impacts AI models. While the strength of AI training data sets is assumed to be their grounding in human data annotation, we show that, in reality, AI trainers attempt to mimic machine annotation to ensure high accuracy rates. We argue this approach could also potentially sabotage the performance of the resulting models. By relentlessly striving for arbitrary measures of "accuracy" and "precision" that dismiss and mishandle human expertise, data companies are, in fact, paradoxically, creating AI models that may be *less* accurate in practice. In this way, performative accuracy has widespread implications when these supposedly accurate models are implemented for their intended tasks, raising important questions about the reliance on and consequences of such models.

Moreover, performative accuracy can serve to **naturalize and legitimize the labor required to meet accuracy standards and the harms they induce**. Actors involved in data production are convinced by and internalize the importance of accuracy for data quality and model development, resulting in AI trainers accepting arbitrary standards as given and unquestionable. This effect is clearly shown in the strong emphasis on achieving high accuracy rates. While a high accuracy rate in data production is meant to validate data quality, any perceived imprecision discredits workers' labor and efforts. Workers are compelled to continuously rework the data until it reaches clients' expectations and meets evolving accuracy criteria, regardless of the excessive labor required for workers, whether accuracy is determined by human ground truth or whether machines play a significant role in determining what is considered "correct." The normalization and legitimization of imposed and constructed accuracy standards, along with the labor required to achieve them, neglect questions regarding whether these standards are reasonable and well justified, or whether the extensive labor required to meet them is necessary. Workers are left in a particularly vulnerable position when performative accuracy is enforced to their detriment. Although they can appeal individual decisions or opt out of specific tasks, AI trainers cannot challenge the sector-wide pursuit of performative accuracy, which is rigidly enforced beyond individual workflows or companies and presented as self-evident and legitimate.

Admittedly, there is a practical need for having accuracy standards, despite their nature as human constructs. However, we would respond by drawing attention to the excessive labor and subsequent harms workers experience when clients set arbitrary standards and request they be strictly implemented, indifferent to the efforts involved and the marginal or indiscernible improvement these extreme standards might bring to data quality or AI models. Investigating the performative aspect of accuracy thus contributes a new lens to understanding the labor and harms induced in the name of accuracy, which we summarize and conceptualize as *precision labor*.

5.3 Conceptualizing Precision Labor

Centering the labor that enacts, performs, and sustains accuracy and precision, we conceptualize a distinct yet under-documented aspect of their work as *precision labor*. This term refers to *the hidden and excessive labor involved in erasing the messy, ambiguous, and uncertain aspects of technology production, to present technology as objective, truthful, and high-quality, even when such pursuit can be excessive, arbitrary, and harmful to laborers.* Precision labor extends previous literature on ghost work [25] and the relational nature of invisible labor in AI production [65, 75] by questioning if excessive labor is well-justified and necessary. Building on research that focuses on the harsh labor conditions in data production [35, 61], precision labor further underscores how these labor practices and their associated harms are normalized and legitimized in the relentless pursuit of performative accuracy. Specifically, we identify three conceptual advances precision labor can offer. These insights deepen our understanding of labor dynamics in AI training and data production and illuminate broader implications for labor practices in other sectors captivated by the allure of high accuracy.

1. Re-centering Arbitrariness in Technical Production. Precision labor provides a lens to re-center the subjectivity, arbitrariness, and uncertainty inherent in technical production, which are rendered obscure in the guise of accuracy. As Grill [28] aptly stated, "what [is accuracy] making in-/visible? ... What accuracy is 'good enough' in what context and for whom?" This perspective reveals the process of establishing, managing, and performing accuracy, thereby revealing the arbitrariness involved in its construction and production. Additionally, it highlights the human labor that is purposefully erased to maintain the guise of trustworthiness, reliability, and certainty in AI systems [28]. While AI models are credited with reducing uncertainty, it is not eliminated but instead transferred to workers. They often bear the burden and undertake substantial physical, emotional, and relational labor in navigating, negotiating, and mitigating uncertainty. Taking Figure 2 as an example, at first glance, the workflow seems robust and logical, offering a framework for "agreement about standards of comparison" and incorporating an "extended network of people" [98]. In practice, however, this workflow often proves to be performative, primarily due to power asymmetries and the top-down decisions imposed on workers. These decisions are consolidated in rounds of revision, resulting in excessive labor, inefficiencies, and harms.

2. Highlighting the Excessive Aspects of Precision Labor. Precision labor reveals that the excessive and hidden labor required to achieve arbitrary and unrealistic accuracy goals is often unnecessary. This insight adds to the existing scholarship on the invisibility and precarity of work in data production and AI training [25, 35, 56]. We highlight the excessive aspects of precision labor in AI training on two main grounds. First, accuracy standards are often shaped by the demands of clients and the requirements of machine actors. In their daily work, workers frequently face pressure to deliver tasks with a unifying accuracy exceeding 95–98%. The justification for having such rigorous and extreme standards is inadequate, raising questions about their necessity. Second, the insistence on high technical accuracy regarding compliance with machine actors can not only disregard workers' expertise and induce excessive labor, but also potentially sabotage the AI models. While accuracy and precision are ideal goals, their related rates and standards must be well-justified and grounded, especially given the excessive labor and harm involved in achieving them.

3. Discussing the Legitimization of Labor and Harm. Precision labor challenges the normalization and legitimization of imposed and arbitrary accuracy standards, and the extensive and hidden labor and emergent harms they induce. Existing work has adequately elaborated on extractive practices and workers' precarious labor conditions in data production [25, 35, 56]. With the concept of precision labor, we show and counteract the normalization and legitimization of such practices in the name of accuracy. Furthermore, drawing from Suchman [83] and Grill [28], we also underscore that meeting high technical accuracy rates does not resolve the persistent uncertainty and ambiguity in how accuracy is defined, established, and measured. Consequently, technical accuracy can be arbitrarily defined by clients but is meticulously enforced by AI trainers. While this enforcement can improve technical outcomes, it simultaneously reinforces the inherent arbitrariness of the criteria, ultimately undermining the very pursuit of accuracy and precision.

Precision labor is particularly relevant in today's technology landscape, given the exponential proliferation of data-driven technologies. Laborers, AI trainers, and beyond are increasingly compelled to comply with the demand for accuracy in machine intelligence. Our research shows that the making of accuracy and precision is performative, and we challenge the legitimacy of arbitrary accuracy goals and the normalization of the labor required to achieve them. We argue that precision labor can be extended beyond data production and AI training to enrich broader theoretical discussions on labor issues in AI production. For instance, in the context of content creation, creators must continually balance the efficiency and accuracy of producing content [50]. Content creators on Reddit and YouTube need to tag their content accurately, aligning with policies and community norms. This balancing act can involve excessive labor but becomes excessive and deeply intertwined with power dynamics and content creators' financial precarity. Importantly, creators' labor and efforts can be overridden by algorithms and users. Yet, such forces tend to be legitimized and normalized in the name of accuracy. Interrogating precision labor prompts a series of important questions: Who defines technical accuracy? Is the accuracy standard reasonable, and how is it justified? Who benefits from the pursuit of accuracy, and who bears the most harm? Can the benefits from such pursuits possibly outweigh the harms caused to workers, and how can we make work in areas such as data production and AI training more ethical and sustainable?

5.4 Moving Forward: Enhancing Transparency and Re-imagining Labor Management and Empowerment in AI Production

Building on prior research and grounded in our empirical findings, in this section, we offer suggestions and alternative approaches for integrating accuracy as a metric into data documentation frameworks to enhance transparency. We advocate for re-imagining labor management practices and fostering collaboration with, as well as empowerment of, local communities in AI production to promote worker agency and well-being.

5.4.1 Include accuracy-related information in documentation frameworks for data and AI production. Previous work has underscored the importance of systematically documenting not only datasets but also data production contexts [4, 17, 24, 32, 34]. Building on this, Díaz et al. [17] proposed extending the documentation framework by including compensations and working conditions. Given the politics and intricate nature of technical accuracy revealed by our findings, we recommend further expanding these documentation frameworks to include accuracy-related information. Specifically, practitioners should treat accuracy as a metric that encompasses more than just unifying and/or context-specific accuracy rates. Documentation should also address the definitions, measurement methods, justification, and managerial processes involved in establishing and enforcing accuracy. Under such frameworks, if clients demand an extremely high accuracy rate, which causes excessive labor and harms, these factors should be accounted for in the compensation structure. Moreover, we stress the urgent need to justify and document projects demanding extremely high accuracy levels while disclosing the evaluation methods, such as the reliance on machine verification systems. By incorporating accuracy information as a central metric, such frameworks can help mitigate excessive labor derived from high accuracy rates, promoting more responsible sourcing practices while enhancing transparency.

5.4.2 Re-imagine labor management practices. Our findings reveal the excessive labor demands and significant harm imposed on workers when the hybrid management pushes *counting regimes* [11] to their extremes. Counting regimes, which "enumerates tasks, errors, accuracy and productivity," often include metrics such as accuracy rates, hourly task counts, and average handling times. While these regimes are commonly used as managerial tools for clients to exercise control and power over data practices and workflows, they often devalue workers' insights and fail to recognize the indispensable role of human discretion in data production and AI training [11]. Echoing Posada [74]'s assertion that "higher quality data requires better working conditions," we advocate for re-imagining labor management practices that not only better account for workers' time and labor but also actively incorporate their expertise.

We further emphasize the critical role of state involvement in improving working conditions. Our empirical findings show that state-led initiatives to establish AI training as a viable career path are insufficient without effective oversight of technology companies' managerial practices and regulations governing working conditions. In the context of China, the widespread acceptance of overwork [33, 44] is exemplified by the "996 work culture" [11] in the technology sector [95] and manifested in the hybrid management in our findings. These practices often normalize excessive workloads and waiting time, devalue human labor, and perpetuate overwork under the guise of prevailing values such as accuracy and flexibility. We urge policymakers to reinforce existing laws to eliminate the residual effects of overwork culture while scrutinizing other problematic managerial techniques, such as hybrid management. In addition, we highly recommend moving away from simplistic metrics such as piecework-based pay or payment determined solely by time spent. Instead, we propose adopting a more holistic approach to payment structure. This approach can align payment with a combination of factors, including overall work time encompassing both structured work hours and time beyond the formal workflow, level of expertise measured by factors (such as years of employment), relevant training and certificates, and productivity. A holistic payment structure could mitigate the harmful effects of precision labor, such as unpaid training and excessive waiting periods. It might also empower workers to assert themselves (e.g., through appeals) and reduce their subordination to machines.

[11] 996 work culture in China refers to managerial requirements, be they explicit or implicit, for workers to work from 9:00 AM to 9:00 PM per day and 6 days per week. The 996 work schedule is no longer legal in China.

5.4.3 Collaborate with and empower local communities. Crucially, we acknowledge that policy and technological solutions alone cannot resolve challenges that are fundamentally social [52]. For instance, our findings show that workers' subordination to machines is shaped by asymmetric power dynamics and prevailing values, such as a preoccupation with technical accuracy, which are further exacerbated by uneven social development (e.g., regional and global socioeconomic disparities) and workers' financial precarity. We recommend methodologies that foreground the community's perspectives and needs, such as community-based research [41], action research [30], and asset-based approaches [23]. For example, projects such as Data Workers' Inquiry [12] and Fairwork [13] build "workplace power" through researchers' support and "shed light on how these technological changes affect working conditions around the world." Building on such collaborative approaches, researchers can integrate community knowledge, demands, and struggles with on-the-ground insights to visualize issues and develop theoretical and analytical frameworks to inform scholarly communities and the public, and support local communities. Such frameworks can improve evidence-based design and policy implications while intervening in socially or technologically deterministic views about and practices around technologies. By fostering collaborations and implementing incremental changes to reduce social disparities, researchers can play a critical role in creating environments where workers enjoy improved career prospects, dignity, and agency.

6 Limitations and Future Work

While our study makes both conceptual and empirical contributions, we acknowledge its limitations. Qualitative and ethnographic methods are powerful because of their situatedness and "interpretative flexibility," enabling researchers to capture the complexity of interlocutors' experiences [72, 80, 82]. As such, we do not claim our findings to be generalizable to data workers and other types of laborers in AI production broadly or in other geographic locations. We encourage future research to employ alternative methods, such as surveys with large representative samples in different settings, including both platform and BPOs environments, to evaluate how our findings might be affirmed, challenged, or complicated. These methods could also extend current work by quantitatively assessing the role of accuracy and precision in shaping worker's experiences. Moreover, precision labor as a lens originates in the context of data work and AI training, but its applicability may extend beyond these areas. Future studies could broaden the scope to explore how precision labor functions in other geographical contexts. For example, as Chinese technology companies expand to South Asian and African countries [14], future work might examine the Chinese presence in these regions, where dominant scholarship has traditionally focused on the influence of Western tech companies. In addition, our fieldwork revealed consistently high accuracy expectations among clients, with 95% being the typical minimum threshold. We encourage future researchers to explore how client budgets influence accuracy requirements and impact workers. For example, scholars might quantitatively investigate the relationship between project budgets, accuracy thresholds, and worker compensation and the associated harms, as well as cases where lower accuracy thresholds might be sufficient. Moreover, while our focus on accuracy emerged inductively as an especially salient theme in our fieldwork data, we also acknowledge AI systems may have other objectives, including latency, fairness, and explainability. Therefore, we encourage future work to explore these objectives holistically concerning AI systems.

[12]https://data-workers.org/
[13]https://fair.work/en/fw/homepage/
[14]2024年，中国企业出海五大趋势)https://finance.sina.com.cn/jjxw/2024-08-06/doc-inchswia4767928.shtml

7 Conclusion

This work has empirically demonstrated how accuracy and precision are perceived, established, and managed in the context of commercial AI training. It also uncovered the harms that come from an obsession with such values. Based on a multi-sited ethnographic study of AI trainers in China, we have analyzed the role of accuracy and precision in workers' everyday work practices and contributed to the organization's perspectives on AI production. Our findings reveal that achieving high technical accuracy often requires workers to undergo extensive training, endure prolonged waiting periods, face financial precarity, and think like machines. To frame interlocutors' experiences, we introduced the concept of *precision labor* as a lens to highlight the excessive, hidden work and the resulting harms (e.g., financial insecurity and human subordination to machines) that emerged from the reckless pursuit of performative accuracy. This research extended the conversation beyond the harms associated with technology implementation to locate and take seriously ethical considerations within the technology production phase [9, 59]. As machines increasingly participate in AI production, our findings reveal how prevailing values such as accuracy and precision are more likely to be used to justify harms and partner with machines to become a new source of epistemic authority. Given the serious negative implications we uncovered in AI production, we proposed suggestions and alternative approaches to promote worker agency and well-being.

Acknowledgments

This research was supported by the International Institute, the Lieberthal-Rogel Center for Chinese Studies, and the Rackham Graduate School at the University of Michigan. We thank the interlocutors in this study for sharing their experiences and anonymous reviewers for their thoughtful comments, which helped improve this work. We are also grateful to Silvia Lindtner and Kentaro Toyama for providing valuable early feedback on this work. Additionally, we appreciate the insights and support from friends and colleagues, including Alexandria Rayburn, Christine Feak, Cami Goray, Cassidy Pyle, David Hartmann, Gabriel Grill, Lu Xian, Pelle Tracey, Qunfang Wu, Renkai Ma, and the members of the AnthroTech Lab and Social Transition Tech Research Group at the University of Michigan. Finally, we thank Emily Albright and Krista Dunger for their graphic design support.

References

[1] Zahra Ashktorab, Michael Desmond, Josh Andres, Michael Muller, Narendra Nath Joshi, Michelle Brachman, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Christine T. Wolf, Evelyn Duesterwald, Casey Dugan, Werner Geyer, and Darrell Reimer. 2021. AI-Assisted Human Labeling: Batching for Efficiency without Overreliance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–27. https://doi.org/10.1145/3449163

[2] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. 2009. Methodologies for data quality assessment and improvement. *Comput. Surveys* 41, 3 (July 2009), 1–52. https://doi.org/10.1145/1541880.1541883

[3] Nancy K Baym. 2015. Connect with your audience! The relational labor of connection. *The communication review* 18, 1 (2015), 14–22.

[4] Emily M. Bender and Batya Friedman. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics* 6 (Dec. 2018), 587–604. https://doi.org/10.1162/tacl_a_00041

[5] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 173–184. https://doi.org/10.1145/3531146.3533083

[6] Judith Butler. 1993. *Bodies that matter: on the discursive limits of "sex"*. Routledge, New York, NY.

[7] Li Cai and Yangyong Zhu. 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal* 14, 0 (May 2015), 2. https://doi.org/10.5334/dsj-2015-002

[8] Scott Allen Cambo and Darren Gergle. 2022. Model Positionality and Computational Reflexivity: Promoting Reflexivity in Data Science. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. https://doi.org/10.1145/3491102.3501998

[9] Antonio A Casilli. 2024. 'End-to-end' ethical AI. Taking into account the social and natural environments of automation. In *Artificial intelligence, labour and society*. ETUI, Brussels, 83–92.

[10] Antonio A Casilli, Paola Tubaro, Clément Le Ludec, and Elinor Wahal. 2019. En la trastienda de la inteligencia artificial. Una investigación sobre las plataformas de micro-trabajo en Francia. *Arxius de Ciències Socials* 41 (2019), 85–108. https://hal-lirmm.ccsd.cnrs.fr/LRI-AO/hal-02554213v1

[11] Srravya Chandhiramowuli, Alex S. Taylor, Sara Heitlinger, and Ding Wang. 2024. Making Data Work Count. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 90 (April 2024), 26 pages. https://doi.org/10.1145/3637367

[12] Kathy Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. SAGE. Google-Books-ID: 2ThdBAAAQBAJ.

[13] Kate Crawford. 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

[14] Meltem Demirkus, James J. Clark, and Tal Arbel. 2014. Robust semi-automatic head pose labeling for real-world face video sequences. *Multimedia Tools and Applications* 70, 1 (May 2014), 495–523. https://doi.org/10.1007/s11042-012-1352-1

[15] Catherine D'ignazio and Lauren F Klein. 2020. *Data feminism*. MIT press.

[16] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. MIT Press, S.l.

[17] Mark Díaz, Ian Kivlichan, Rachel Rosen, Dylan Baker, Razvan Amironesei, Vinodkumar Prabhakaran, and Emily Denton. 2022. CrowdWorkSheets: Accounting for Individual and Collective Identities Underlying Crowdsourced Dataset Annotation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 2342–2351. https://doi.org/10.1145/3531146.3534647

[18] Lisa Ehrlinger and Wolfram Wöß. 2022. A Survey of Data Quality Measurement and Monitoring Tools. *Frontiers in Big Data* 5 (March 2022), 850611. https://doi.org/10.3389/fdata.2022.850611

[19] Hamid Ekbia and Bonnie Nardi. 2014. Heteromation and its (dis) contents: The invisible division of labor between humans and machines. *First Monday* (2014).

[20] Hamid R Ekbia and Bonnie A Nardi. 2017. *Heteromation, and other stories of computing and capitalism*. MIT Press.

[21] Robert M Emerson, Rachel I Fretz, and Linda L Shaw. 2011. *Writing ethnographic fieldnotes*. University of Chicago press.

[22] Melanie Feinberg. 2017. A Design Perspective on Data. In *Proc. CHI 2017*. ACM, Denver Colorado USA, 2952–2963.

[23] Aakash Gautam, Deborah Tatar, and Steve Harrison. 2020. Crafting, Communality, and Computing: Building on Existing Strengths To Support a Vulnerable Population. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376647

[24] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (Dec. 2021), 86–92. https://doi.org/10.1145/3458723

[25] Mary L. Gray and Siddharth Suri. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt, Boston.

[26] Mary L. Gray, Siddharth Suri, Syed Shoaib Ali, and Deepti Kulkarni. 2016. The Crowd is a Collaborative Network. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, San Francisco California USA, 134–147. https://doi.org/10.1145/2818048.2819942

[27] Ben Green. 2020. The false promise of risk assessments: epistemic reform and the limits of fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 594–606. https://doi.org/10.1145/3351095.3372869

[28] Gabriel Grill. 2022. Constructing Certainty in Machine Learning: On the performativity of testing and its hold on the future. (2022).

[29] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. https://doi.org/10.1145/3173574.3174023

[30] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Trans. Comput.-Hum. Interact.* 18, 3, Article 15 (Aug. 2011), 20 pages. https://doi.org/10.1145/1993060.1993065

[31] Arlie Russell Hochschild. 2022. The managed heart. In *Working In America*. Routledge, 40–48.

[32] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2018. The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards. http://arxiv.org/abs/1805.03677 arXiv:1805.03677 [cs].

[33] Hongwei Huang, Xue Xia, Wenxi Zhao, Xiaoqing Pan, and Xiaoqiang Zhou. 2021. Overwork, job embeddedness and turnover intention among Chinese knowledge workers. *Asia Pacific Journal of Human Resources* 59, 3 (2021), 442–459.

[34] Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. 2021. Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 560–575. https://doi.org/10.1145/3442188.3445918

[35] Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 611–620. https://doi.org/10.1145/2470654.2470742

[36] Steven J Jackson. 2014. 11 rethinking repair. *Media technologies: Essays on communication, materiality, and society* (2014), 221–39.

[37] Ju Yeon Jung, Tom Steinberger, John L. King, and Mark S. Ackerman. 2022. How Domain Experts Work with Data: Situating Data Science in the Practices and Settings of Craftwork. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–29. https://doi.org/10.1145/3512905

[38] Sean Kross and Philip Guo. 2021. Orienting, Framing, Bridging, Magic, and Counseling: How Data Scientists Navigate the Outer Loop of Client Collaborations in Industry and Academia. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–28. https://doi.org/10.1145/3476052

[39] Thomas S. Kuhn. 1977. *The essential tension: selected studies in scientific tradition and change*. University of Chicago Press, Chicago.

[40] Maurizio Lazzarato. 1996. Immaterial labor. *Radical thought in Italy: A potential politics* 1996 (1996), 133–47.

[41] Christopher A. Le Dantec and Sarah Fox. 2015. Strangers at the Gate: Gaining Access, Building Rapport, and Co-Constructing Community-Based Research. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1348–1358. https://doi.org/10.1145/2675133.2675147

[42] Clément Le Ludec, Maxime Cornet, and Antonio A Casilli. 2023. The problem with annotation. Human labour and outsourcing between France and Madagascar. *Big Data & Society* 10, 2 (July 2023), 20539517231188723.

[43] Kai-Fu Lee. 2018. *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin.

[44] Xiaotian Li. 2023. Managerial technique and worker subjectivity in dialogue: Understanding overwork in China's internet industry. *Work, Employment and Society* 37, 6 (2023), 1699–1716.

[45] Cindy Lin and Silvia Margot Lindtner. 2021. Techniques of Use: Confronting Value Systems of Productivity, Progress, and Usefulness in Computing and Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445237

[46] Cindy Kaiying Lin and Steven J Jackson. 2023. From Bias to Repair: Error as a Site of Collaboration and Negotiation in Applied Data Science Work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.

[47] Silvia M Lindtner. 2020. *Prototype nation: China and the contested promise of innovation*. Vol. 29. Princeton University Press.

[48] Helen E. Longino. 1995. Gender, Politics, and the Theoretical Virtues. *Synthese* 104, 3 (Sept. 1995), 383–397.

[49] Helen E. Longino. 1996. Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy. In *Feminism, Science, and the Philosophy of Science*, Lynn Hankinson Nelson and Jack Nelson (Eds.). Springer Netherlands, Dordrecht, 39–58. https://doi.org/10.1007/978-94-009-1742-2_3

[50] Renkai Ma, Zinan Zhang, Xinning Gui, and Yubo Kou. 2024. Labeling in the Dark: Exploring Content Creators' and Consumers' Experiences with Content Classification for Child Safety on YouTube. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 1518–1532.

[51] Donald MacKenzie. 1993. *Inventing accuracy: A historical sociology of nuclear missile guidance*. MIT press.

[52] Mirca Madianou. 2014. Smartphones as polymedia. *Journal of Computer-Mediated Communication* 19, 3 (2014), 667–680.
[53] George E Marcus. 1995. Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual review of anthropology* 24, 1 (1995), 95–117.
[54] David Martin, Benjamin V. Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, Baltimore Maryland USA, 224–235. https://doi.org/10.1145/2531602.2531663
[55] Colten Meisner, Brooke Erin Duffy, and Malte Ziewitz. 2024. The labor of search engine evaluation: Making algorithms more human or humans more algorithmic? *New Media & Society* 26, 2 (2024), 1018–1033.
[56] Milagros Miceli and Julian Posada. 2022. The Data-Production Dispositif. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–37. https://doi.org/10.1145/3555561
[57] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–25. https://doi.org/10.1145/3415186
[58] Ministry of Human Resources and Social Security, Ministry of Industry and Information Technology of the People's Republic China. 2021. National Occupational Skill Standard Occupational Code: 4-04-05-05 Artificial Intelligence Trainers (2021). Accessed by March 3, 2024.
[59] James Muldoon, Callum Cant, Mark Graham, and Funda Ustek Spilda. 2023. The poverty of ethical AI: impact sourcing and AI supply chains. *AI & SOCIETY* (Dec. 2023). https://doi.org/10.1007/s00146-023-01824-9
[60] James Muldoon, Callum Cant, Boxi Wu, and Mark Graham. 2024. A typology of artificial intelligence data work. *Big Data & Society* 11, 1 (March 2024), 20539517241232632. https://doi.org/10.1177/20539517241232632
[61] James Muldoon, Mark Graham, and Callum Cant. 2024. *Feeding the Machine: The Hidden Human Labor Powering A.I.* Canongate Books, Edinburgh.
[62] Michael Muller, Cecilia Aragon, Shion Guha, Marina Kogan, Gina Neff, Cathrine Seidelin, Katie Shilton, and Anissa Tanweer. 2020. Interrogating Data Science. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. ACM, Virtual Event USA, 467–473. https://doi.org/10.1145/3406865.3418584
[63] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Glasgow, Scotland Uk, 1–15.
[64] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, and Casey Dugan. 2021. Designing Ground Truth and the Social Life of Labels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445402
[65] Gemma Newlands. 2021. Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data & Society* 8, 1 (Jan. 2021), 205395172110160. https://doi.org/10.1177/20539517211016026
[66] Jack E. Olson. 2003. *Data quality: the accuracy dimension*. Morgan Kaufmann, San Francisco.
[67] Samir Passi and Steven Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 2436–2447. https://doi.org/10.1145/2998181.2998331
[68] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 1–28. https://doi.org/10.1145/3274405
[69] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (July 2020), 205395172093960. https://doi.org/10.1177/2053951720939605
[70] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (Nov. 2021), 100336. https://doi.org/10.1016/j.patter.2021.100336
[71] Alan Peshkin. 1988. In search of subjectivity—one's own. *Educational researcher* 17, 7 (1988), 17–21.
[72] Trevor J. Pinch and Wiebe E. Bijker. 1984. The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science* 14, 3 (Aug. 1984), 399–441. https://doi.org/10.1177/030631284014003004 Publisher: SAGE Publications Ltd.
[73] Julian Posada. 2022. Embedded reproduction in platform data work. *Information, Communication & Society* 25, 6 (2022), 816–834.
[74] Julian Posada. 2023. Platform Authority and Data Quality: Who Decides What Counts in Data Production for Artificial Intelligence? In *Decoding Digital Authoritarianism*, Ziyaad Bhorat and Martin Rauchbauer (Eds.). Berggreun Institute, Los Angeles, CA, 81–85. https://berggruen.org/news/decoding-digital-authoritarianism
[75] Noopur Raval. 2021. Interrupting invisibility in a global world. *Interactions* 28, 4 (June 2021), 27–31. https://doi.org/10.1145/3469257
[76] Annabel Rothschild, Amanda Meng, Carl DiSalvo, Britney Johnson, Ben Rydal Shapiro, and Betsy DiSalvo. 2022. Interrogating Data Work as a Community of Practice. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–28. https://doi.org/10.1145/3555198
[77] Shruti Sannon and Dan Cosley. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300512
[78] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–37. https://doi.org/10.1145/3476058
[79] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2018. Data Work in a Knowledge-Broker Organisation: How Cross-Organisational Data Maintenance Shapes Human Data Interactions. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference* (Belfast, United Kingdom) *(HCI '18)*. BCS Learning & Development Ltd., Swindon, GBR, Article 14, 12 pages. https://doi.org/10.14236/ewic/HCI2018.14
[80] Mario Luis Small. 2009. How many cases do I need?' On science and the logic of case selection in field-based research. *Ethnography* 10, 1 (2009), 5–38.
[81] Luke Stark and Jesse Hoey. 2021. The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 782–793. https://doi.org/10.1145/3442188.3445939
[82] Julia C. Strauss. 2013. China and Africa Rebooted: Globalization(s), Simplification(s), and Cross-cutting Dynamics in "South–South" Relations. *African Studies Review* 56, 1 (April 2013), 155–170. https://doi.org/10.1017/asr.2013.10 Publisher: Cambridge University Press.
[83] Lucy Suchman. 2020. Algorithmic warfare and the reinvention of accuracy. *Critical Studies on Security* 8, 2 (2020), 175–187.
[84] TL Taylor, Tom Boellstorff, Bonnie Nardi, and Celia Pearce. 2013. *Ethnography and virtual worlds: A handbook of method*. Princeton university press.
[85] Andrea K. Thomer, Dharma Akmon, Jeremy J. York, Allison R. B. Tyler, Faye Polasek, Sara Lafia, Libby Hemphill, and Elizabeth Yakel. 2022. The Craft and Coordination of Data Curation: Complicating Workflow Views of Data Science. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–29. https://doi.org/10.1145/3555139
[86] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the Invisible Labor in Crowd Work. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–26. https://doi.org/10.1145/3476060
[87] Anna Lowenhaupt Tsing. 2012. On nonscalability: The living world is not amenable to precision-nested scales. *Common knowledge* 18, 3 (2012), 505–524.
[88] Paola Tubaro and Antonio A Casilli. 2019. Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics* 46 (2019), 333–345.
[89] Paola Tubaro, Antonio A Casilli, and Marion Coville. 2020. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society* 7, 1 (Jan. 2020), 2053951720919776. https://doi.org/10.1177/2053951720919776 Publisher: SAGE Publications Ltd.
[90] Matheus Viana Braz, Paola Tubaro, and Antonio A. Casilli. 2023. *Microwork in Brazil. Who are the workers behind artificial intelligence?* Research Report. DiPLab & LATRAPS. https://diplab.eu/?p=2833
[91] Yair Wand and Richard Y. Wang. 1996. Anchoring data quality dimensions in ontological foundations. *Commun. ACM* 39, 11 (Nov. 1996), 86–95. https://doi.org/10.1145/240455.240479
[92] Dakuo Wang, Josh Andres, Justin D. Weisz, Erick Oduor, and Casey Dugan. 2021. AutoDS: Towards Human-Centered Automation of Data Science. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–12. https://doi.org/10.1145/3411764.3445526
[93] Ding Wang, Shantanu Prabhat, and Nithya Sambasivan. 2022. Whose AI Dream? In search of the aspiration in data annotation.. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. https://doi.org/10.1145/3491102.3502121
[94] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. https://doi.org/10.1145/3359313
[95] Jenny Jing Wang. 2020. How managers use culture and controls to impose a '996'work regime in China that constitutes modern slavery. *Accounting & Finance* 60, 4 (2020), 4331–4359.
[96] R.Y. Wang, V.C. Storey, and C.P. Firth. 1995. A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering* 7, 4 (Aug. 1995), 623–640. https://doi.org/10.1109/69.404034

[97] Mark E. Whiting, Dilrukshi Gamage, Snehalkumar (Neil) S. Gaikwad, Aaron Gilbee, Shirish Goyal, Alipta Ballav, Dinesh Majeti, Nalin Chhibber, Angela Richmond-Fuller, Freddie Vargus, Tejas Seshadri Sarma, Varshine Chandrakanthan, Teogenes Moura, Mohamed Hashim Salih, Gabriel Bayomi Tinoco Kalejaiye, Adam Ginzberg, Catherine A. Mullings, Yoni Dayan, Kristy Milland, Henrique Orefice, Jeff Regino, Sayna Parsi, Kunz Mainali, Vibhor Sehgal, Sekandar Matin, Akshansh Sinha, Rajan Vaish, and Michael S. Bernstein. 2017. Crowd Guilds: Worker-led Reputation and Feedback on Crowdsourcing Platforms. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 1902–1913. https://doi.org/10.1145/2998181.2998234

[98] M. Norton Wise (Ed.). 1995. *The values of precision.* Princeton Univ. Press, Princeton, NJ.

[99] Di Wu. 2023. Good for tech: Disability expertise and labor in China's artificial intelligence sector. *First Monday* (Jan. 2023). https://doi.org/10.5210/fm.v28i1.12887

[100] Tianling Yang and Milagros Miceli. 2024. "Guilds" as Worker Empowerment and Control in a Chinese Data Work Platform. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 365 (Nov. 2024), 27 pages. https://doi.org/10.1145/3686904

[101] Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 1–23. https://doi.org/10.1145/3392826