Match Made by Humans: A Critical Enquiry into Human-Machine Configurations in Data Labelling

Srravya Chandhiramowuli
International Institute of Information Technology Bangalore
srravya.c@gmail.com

Bidisha Chaudhuri
International Institute of Information Technology Bangalore
Bidisha@iiitb.ac.in

Abstract

We present a critical ethnographic study of data labelling conducted in a Bangalore-based AI start-up. Labelled datasets are primarily produced by human data workers. We explore how humans and machines are configured together in data labelling and what are the demands placed on human workers, including on their body and cognition, while being assigned in the service of machine intelligence. We also show how these human-machine configurations sustain and reproduce the seamless functioning of apparently "autonomous" AI as a normative vision. Though labelled datasets are an indispensable prerequisite to creating ML/AI-based systems, the human labour that produces these datasets cannot be acknowledged fully if the techno-entrepreneurial vision of "self-learning" machine intelligence is to be celebrated and sustained. In pursuit of a normative position of what AI should be, we are left with a denial of how AI is actually produced now.

Keywords: data labelling, AI, labour, work practices,

1. Introduction

Most of the recent literature on Artificial Intelligence (AI) and human labour focuses on different ways in which deployment of AI in labour processes impacts human workers and their prospects of employment (Posada, 2020). One set of literature in this category focuses on the significance of human employment in face of mass deployment of AI at work (Frey & Osborne, 2017), new types of work created for human workers (Autor, 2015), the process of recruitment (Ajunwa & Greene, 2019), and implications for agency, privacy and surveillance of human workers (Adams-Prassl, 2020; Lee et al., 2015; Moore & Robinson, 2016). Another line of enquiry examines the restructuring of labour processes through deployment of AI at workplaces, such as emergence of gig work (Woodcock & Graham, 2020), online freelancing (Wood et al., 2018), ghost work (Gray & Suri, 2019), and microwork (Irani, 2015a). Though much of this scholarship unravels the emerging trends in work and employment in current times, they are generally considered to be signalling towards a new paradigm of "future of work" (Posada, 2020) being produced through the deployment of AI at the workplace.

In this emerging body of scholarship on AI and human labour, an important aspect that remains relatively understudied but slowly gaining currency, is the role of human labour in the development and maintenance of AI (Gray & Suri, 2019; Miceli et al., 2020; Sambasivan et al., 2021). Joyce et al (2021) call for sociological focus on the production of AI as sociotechnical systems so that we may better understand AI's division of labour and the significance of organisational contexts and labour practices in shaping AI systems. Rather than theorise about AI's impact on the future of work, this line of enquiry pays specific attention to human-machine configurations in the production of AI, here and now (Crawford & Joler, 2018; Posada, 2020). We draw on this emerging body of literature, in this paper, to explore how humans and machines are configured together in the production of machine intelligence and what are the demands placed on the human worker, including on their body and cognition, while being assigned in the service of machine intelligence. Finally, we show how these human-machine configurations sustain and reproduce the seamless functioning of apparently "autonomous" AI as a normative vision.

In doing so, we engage in a critical study of the human data labelling, focussing on its specific work arrangements and labour processes. We find that data labelling by human workers not only supplements the machine's ability but also fulfils techno-entrepreneurial visions of machine intelligence that tend to undermine human labour in favour of the machine. Examining this work arrangement that builds, sustains and makes AI

URI: https://hdl.handle.net/10125/102882
978-0-9981331-6-4

possible, here and now, is crucial in advancing our collective understanding of the future of technology-mediated work and its impact on human workers.

We present our work through the following four sections. We provide a brief review of relevant work around AI, data and human work and how that shapes our enquiry of data labelling. In the following section, we explain our methodological approach that includes our choice of methods, description of field sites and methods of data analysis. We then present our findings about the human-machine configurations within data labelling by explicating the work practices of the individual data labellers, the data labelling team as a whole and the organisational dynamics between data labelling team, the engineering teams and the company management. In the next section we situate these findings within the broader discussion of AI and human labour and its implications for human workers. Finally, we conclude by pointing out the ways in which our understanding of human-machine configuration in data labelling may contribute towards new ways of discussing AI ethics that look beyond deployment of AI and its future implications and focus more on the current conditions of AI production, paying critical attention to structural conditions embedded in its labour processes.

2. Relevant Work

Over the last decade, there has been an accelerated development of algorithmic capabilities, particularly in machine learning, artificial intelligence, big data and the internet of things. These technologies generate, extract, and synthesise large datasets, harnessing them for insights that can be commodified. The datasets used by these systems require significant cleaning, classification, and verification; this is primarily dependent on human effort (Bilić, 2016; Sambasivan et al., 2021). However, the technology industry masks this human effort in data work by foregrounding the algorithmic capabilities in ways that hide the data tasks necessary to build, train and support algorithms (Gray & Suri, 2019; Irani, 2015a; Roberts, 2016; Shestakofsky, 2017). Crawford and Joler (2018) in their study of Alexa, the virtual assistant developed by Amazon, highlight the different types of human labour (of engineers and technicians and "crowd-workers") that go into the creation, operation and sustenance of the autonomous system and device. Others focus on data work within the AI ecosystem that often gets side-lined as mundane and non-cognitive in comparison to "creativity" of model work. They foreground the relevance, mundaneness, subjectivities and complexities of data work in automated technological systems to underline the importance of data work in development of AI. (Gray & Suri, 2019; Miceli et al., 2020; Sambasivan et al., 2021). Besides highlighting the value of data labelling work, they also show how systematic neglect of this human work leads to further exploitation and displacement of those doing this work (Irani, 2015b), relegating them to a spiralling cycle of invisibility and irrelevance amidst celebratory production of AI. This creates a smokescreen around how humans and machines are configured in the service of automation. Against this backdrop, we aim to elaborate the work practices involved in data labelling through our ethnographic study in a data labelling team in a start-up firm in Bangalore that produces AI-based price intelligence products.

Furthermore, we place our understanding of these work practices in conversation with the literature on the political economy of automation and human labour (Ekbia & Nardi, 2017; Poster et al., 2016; Raval, 2021) that informs our work about the structural inequalities of the AI industry. Much like the outsourcing industry of the early information technology (IT) services, data work in the contemporary AI industry is located at the bottom of the global value chain of AI production, displaced away from the centres and processes of model development, either by gig work/crowd work model or through outsourcing to Global South (Graham et al., 2017; Pietrobelli & Rabellotti, 2011). The AI/ML companies of the Global North employ data workers from lower-income countries to support their data production functions (Miceli & Posada, 2022; Murali, 2019). While on one hand, these jobs are celebrated in the Global South for their ability to provide livelihood opportunities to a large population of educated youth, (Graham et al., 2017; Joshi, 2019; Murgia, 2019), on the other hand, they are accompanied by perceptions of being 'low-skill' jobs. Thus, data work becomes a new site for devaluing human labour and skills within the larger discourse of automation (Ekbia and Nardi, 2017).

The data labelling workforce is kept apart from the rest of the AI production process, trained mainly to obey their requesters (Miceli & Posada, 2022). They are undervalued and neglected within the technology industry as their jobs form the 'mundane', 'repetitive' and 'non-cognitive' aspects of developing

AI. In her study of platform-based crowdsourced work, Irani (2015a) notes that this serves to sustain the difference between "innovative" work and "menial" labour in high-technology work. The distanced crowdworkers on platforms like Amazon Mechanical Turk produced surplus value for their requesters but it was their invisibility that fuelled the status of the companies that employed them (Irani, 2015c). Similarly, Gupta's (2019) study illustrates the role of feminised precarious digital labour in enabling postcolonial technocapitalism to flourish in the Indian startup ecosystem. Shestakofsky's (2017) longitudinal study of "human-software complementarity" shows that organisations "continually reconfigure assemblages of software and human helpers", as they adapt their goals in response to the pressures and compulsions of venture capital funding. Taken together, these studies underscore the importance of examining human-machine configurations in sociotechnical systems. We build on this body of work to analyse how the human labour of data labelling was organised and configured in the service of machine intelligence at an AI startup in Bangalore, India.

3. Methods

We explored the work arrangements and labour processes of data labelling in a critical qualitative study of data labelling work in India. Owing to its large population of educated youth and availability of digital infrastructures, India is an important site in the Global South for outsourced or crowdsourced data work. We conducted an exploratory, inductive enquiry into data labelling work at a start-up that builds AI-based pricing analytics for global e-commerce platforms and brands. The company used computer vision and deep learning models to determine if two products from two competing e-commerce platforms were the same, so as to compare their prices, offers and discounts and provide pricing analytics. This AI-led product matching was supported by the data labelling team that verified the automated matches made by the model as well as manually searched and added matches for products that the model could not. The in-house data labelling team was called the QA team and its members, QA analysts. It was a 30-member team that was headed by the QA manager and divided into four smaller teams, each led by a team-lead.

We adopted ethnographic methods of participant observation and in-depth interviews to closely examine the organisational structures and labour processes that shape data labelling. The first author joined the QA team as an intern for six weeks between March - May 2021. During this period, we conducted 20 in-depth interviews with QA analysts, team-leads, manager, developers as well as the company's top management. Due to Covid-19 restrictions at the time of the study and the organisation's work from home policies, the participant observation and most of the interviews were conducted online. This meant that there were limited opportunities to join or observe social, workplace interactions that form an integral part of ethnographic studies of the workplace. In the virtual environment, the first author relied on following the internal group chat that the QA team used to pick up on team chatter and activities, and later followed up on them with individual team members in direct messages and scheduled interviews. Our enquiry was guided by a constructivist grounded theory approach (Charmaz, 2017), which allowed us to examine emergent job roles in data labelling and their associated work practices in an indeterminate, but reflexive manner.

In the findings section that follows, we present vignettes drawn from the experiences and observations of the first author as a data labelling intern. Hence the findings section is written as a first-person account of the participant observer. We also use this as a way of making the researcher's presence in the context visible. We use pseudonyms for the company (referring to it as PriceWise) as well as the interviewees to protect their identity.

4. Findings

4.1. The human labour of product matching

On the first day of my internship in PriceWise's QA team, Jai, one of the senior QA analysts, gave me a demo of the browser-based QA tool, designed and developed in-house at PriceWise, specifically for product matching and verification tasks. For each product, the tool would display a title, thumbnail image of the product, its price, availability, and a list of suggested matches. The suggested matches were often similar to the base product but they could differ in colour, size, gender or age category, or even model. Suggested matches were products identified as matches by the computer vision algorithm but with a low confidence score. The task at hand for me and any QA analyst was to find a valid match for the base product from the list of suggested matches.

After the demo of the QA tool, Jai assigned 20 sample products for me to match. Most new joinees like me would be assigned fashion brands to begin with, as those were considered easier than other product categories. On that first day, I clicked, opened, read the product details and looked for matches in the suggested matches list. But I struggled to find exact matches for the products assigned to me. Firstly, the products and brands were unfamiliar to me; I had never heard of most of them until then. PriceWise's clients were predominantly based in North America; their e-commerce products and brands that we matched were either unavailable or classified as luxury goods in India. Therefore, their names were not common knowledge in India. Further, I found that the colour of the same pair of pants that was "Navy" on one platform, was "Dark Blue/Denim" on another; the colour of a handbag was listed as "Milk" on one site and "Snow" on another. In both cases, I could see from the product images that the base product and its match were of the same colour, yet the colour-names provided were different. Similarly, I also found that product names and descriptions also varied across platforms. I once found the exact same "fanny pack" listed as a "belt bag" on a competitor platform. It was up to me, the human labeller, to determine whether these products were a match or not, despite some apparent differences in certain details. On that first day, I failed to match even a single product but got a real taste for what product matching entailed!

Over the next couple of weeks, as I gained confidence in matching fashion products, my team-lead started assigning other kinds of client accounts, those involving grocery and industrial equipment. For grocery products, I had to pay attention to quantity, dimensions, and volume. Unlike in fashion, I could no longer rely on the product images to spot the differences, since the same soap wrapper image would be provided whether it was a 3-pack or a 6-pack. Besides finding matches, I would also be assigned the work of verifying algorithmic matches. This involved checking whether the computer vision algorithm had indeed chosen the right match for the base product. I found this work to be far easier than searching for matches, since it only involved comparing two products with each other. As I found and verified more product matches and across different clients, my eyes became well-trained to scan the base product title for keywords including colour, gender, category (such as tall, plus, toddler, etc.), size, quantity (pack of 2 or 4), and any specific features like a puff sleeve, or lemon-flavoured noodles.

As I paid attention to these details, I also completely skipped some others like product price or availability; those were not important to consider while matching. The price comparison would be done algorithmically once we identified and verified matches. The job of an QA analyst was limited to product matching, with little to no insight on how the data came into the tool or what happened to it when it left the tool. This was most evident in the pre-QA process. That was an internal QA process initiated whenever a new client was onboarded, to test if the existing algorithmic system worked well for their products. The QA team would take a sample of about 20-50 of the client's products and try to find matches for them. They would then record whether the suggested matches listed any relevant matches using the Feedback button.

In my first week as an intern, I was curious to try out the Feedback button on the top right corner of the products page. When I clicked on it, a pop-up box appeared with the title "Issues detected" and about a dozen options listed under it like "Correct match unavailable", "Correct match appears in first 10 suggestions", "Similar product with different colour", "Different pack size/quantity", "Wrong product type appearing in suggestions" and such. I asked Jai if I should give feedback for each product. He told me it was not needed; feedback was required only during "pre-QA". The following week I was assigned some products from a new client account for pre-QA matching. It was like the regular product matching, with the additional step of submitting feedback for each product. And when the tool did not provide a relevant match under the suggested matches list, I had to manually (outside the tool) go to the competitor's website, search for the product and if available, add its link in the feedback box. A few days later, when I was assigned the same client account for regular QA matching, I found that the tool now listed relevant matches in the suggestions list for most products. I was curious how the tool was now able to display relevant suggestions which it did not earlier. I wondered if it was the feedback we provided at the pre-QA stage.

I asked my peers in the team if they knew how our feedback was used by the tool but they were not sure themselves - some said team-leads looked at it to understand the client, some others felt it went to the computer vision team to tweak the model for this client and a few believed it was only for internal documentation. It was only the QA Manager who could confidently confirm that it was indeed for the computer vision team to adjust the algorithmic system's performance for the particular client. Though almost everyone in the team had been giving pre-QA feedback regularly, they were in the dark about how their feedback helped the computer vision team to improve the model. They viewed the tool as a site for performing their daily tasks but rarely did they see their work as also transforming the tool itself.

4.2. Mentorship for accuracy, competition for productivity

Every evening, Rishav would ping me on Hangouts at around 5PM. He was a senior analyst like Jai and responsible for reverifying the work I did every day. This was the team's practice whenever a newcomer joined them. Until the new joinee became confident with QA tasks, senior analysts would re-check their matches everyday and correct any mistakes found in their work.

Rishav: Could you open product ID #3006?

Me: Yes, I have opened it…

Rishav: Can you identify what is the mistake here?

Me: Both the watches are the same right? I have checked the model, brand, dial colour, everything…

Rishav: If you see the seconds needle in the watch, the colour is different… in the base product, it is yellow and in the match you have added, it is white.

Me: Oh, I didn't notice that. Do we have to look at all that also?

Rishav: Yes, for example, in ripped jeans, where is the tear? You may find a match with the exact same brand, colour, size but the tear may be below the knee and in your base product, it may be above the knee. Or for a leather shoe, whether it is a matte finish or gloss finish. All these details matter.

Me: I had no idea that I should look this closely…

Rishav: No issue, you are just starting to learn right… that's why I am telling you what the corrections are, so next time you will be careful. You deleted the wrong match, right?

Me: Yes, I deleted…

Rishav: Thank you

These daily conversations which took a lot of time and energy on Rishav's part were considered crucial for new joinees like me to learn and become better at our work. More importantly, it was essential to ensure high accuracy of the team's overall output. Besides reverifying my work, senior QA analysts like Rishav and Jai were also involved in Quality Control or QC, another QA work practice aimed at reducing the errors in the team's output. The QA tool had a feature to sort completed products by the price difference between the base product and its match. One evening, Rishav shared his screen over a video call to show how he used this feature to identify product matches with a significant price difference (up to 40%) and re-verify those matches. He told me that this practice was introduced in response to client escalations about incorrect products being compared for pricing analytics. The QA team was expected to maintain an accuracy of at least 95% in their matching work and mentoring and reverification practices were fashioned to ensure the same.

Accuracy was not the only priority for the team; they also prioritised high productivity in order to meet client delivery deadlines. Just as they adopted mentoring as an approach to achieve high accuracy, the QA team fostered a sense of competition among its analysts to maintain high productivity.

In the first week of April, I received an email with the subject line, "Monthly Stats - April 2021", sent out to the whole team by one of the team-leads, Swetha. In this email, came attached a spreadsheet which listed all the members of the QA team in a table, with some names in red and green. On closer look, I saw that it was a ranked list of QA analysts, based on their performance in March 2021. It listed the names of the analysts, the number of hours they worked in March, the number of products they matched in that period (referred to as productivity), the percentage of error in their work, a weighted score calculated based on their productivity and error rate, and a corresponding rank reflecting their position in the table. Additionally, the first three names in the list were labelled in green as Top 1, Top 2 and Top 3, a handful of names that followed them were labelled "ABOVE AVERAGE", following them were a few "BELOW AVERAGE", finally the last 3 names in the list were marked in red and labelled Bottom 3, Bottom 2, Bottom 1. This was a monthly exercise of evaluating and ranking QA analysts by counting, aggregating, and comparing their work with the rest of the team.

Priya, a QA analyst who featured in the fifth position that month, viewed monthly stats as a motivation to keep improving her performance.

That's why we are having monthly stats, so that we do more! It is good actually. It makes us more competitive, to do more and our productivity will also go up. If we come in the top 3, we even have cash rewards!

- Priya, QA analyst

Besides the cash rewards, analysts were evaluated for appraisals and promotions on the basis of the monthly stats. Consequently, they found novel ways to boost their stats, Pushkar, one of the QA team leads, noted.

What QA people will do is… If there is a difficult account, they would work normal hours… like 7-8 hours and close their laptop. If there is an easy account right, they will work 9-10 hours, increasing 1-2 hours extra on those days. This will increase their stats count, and they can come in the top or above average. Anyway, the stats will be calculated on a monthly basis, not daily basis.

- Pushkar, QA team lead

While the analysts chased their stats and ranking, team leads often reminded them that accuracy was still the first priority.

To my team, I always say, please don't look at your productivity, only focus on your accuracy! Accuracy is more important than productivity. That is the reason humans are doing the QA part, right? Otherwise, machines can just quickly do some matches, but that's not accurate. That's where we come in. So, first focus on accuracy.

- Pushkar, QA team lead

Though accuracy was the priority for the QA team's work, by quantifying everyday work into statistics and tying those statistics to performance evaluation and rewards, QA analysts were left to juggle both, accuracy, and productivity. They could not afford to drop either metric, for one was critical to the machine's performance and the other was critical to their own future in the team and company.

4.3. Techno-entrepreneurial notions of data labelling

It was the 10th anniversary celebration at PriceWise on April 1st, 2021. Samantha, the QA manager, graciously invited me to join the celebratory event that was happening online, in wake of Covid-19. In the hour-long celebration, Naraen, the company's CTO and co-founder, eloquently recounted the story of onboarding their very first client (a children's diaper company), securing their initial round of funding, and hiring Vivek, a computer architect who designed the pricing analytics pipeline and Sampath, a senior AI engineer who was instrumental in building the company's computer vision-based product matching algorithms. As I listened to his speech, it was remarkable that the QA team found no mention in his speech. The QA team had been a part of the company since its inception, integral to delivering pricing accurate analytics, indispensable to the AI development pipeline, and one of the largest teams in the company. Yet their work or members did not receive any special mention, recognizing their contribution to the company's success. Perhaps, it was because the company's top management viewed, as we learnt through our interviews with them, QA as simple work that "anybody could do".

All these e-commerce products are in the public domain, you can Google and figure out what they are. We just have to train them on the [QA] tool. Then… they need to be trained on product features, attributes, and categories. For example, the spec [specification] for mobile phones or washing machines. This is very general knowledge, right? Anyone can pick it up from the web also.

- Top-level executive at PriceWise

The product matching even any normal graduate can do, you don't need any technical knowledge to do that. You only have to look at the attributes that will be told to you and then do the matching

- Samantha, QA Manager

Though we repeatedly heard in our interviews that QA work could be done by anybody with basic digital literacy and online shopping exposure, I found myself struggling to cope with the day-to-day demands of QA work. The initial challenge lay in learning the many nuances of product matching. Then, as the team-leads began to assign regular work to me, it involved looking at the computer screen continuously and meticulously for at least 6-8 hours a day that my eyes felt the strain for the first week or so. We needed to be able to switch context across the different categories of products, different clients, and keep up with changing instructions. On the same day, we could be matching clothing products where we could rely on product images and then grocery products where we could not rely on images and later switch to jewellery, where we might have to count the number of stones in an earring before matching them. Even as I managed to complete these tasks without errors, I struggled to consistently deliver high productivity, matching products day after day, without making mistakes and without slowing down. None of these demanding aspects of the QA analyst's role were captured when PriceWise's top executives described the work as something "anybody could do".

PriceWise's leadership consisted of engineers and computer scientists who strived to develop scalable, AI-driven data analytics. Their core expertise lay in using data and machine learning technologies to build pricing analytics systems. They recognised early-on in their entrepreneurial journey that they needed human labellers to deliver accurate analytics if they were to stay competitive in the pricing intelligence domain. Though they constituted the QA team for this purpose, their technological training led them to view this human involvement as a "sub-optimal" approach.

> *See we are all techies, and we know this is not the right way of doing it. from a technical point of view… this is a sub-optimal and inefficient way. Basically, the engineer in you is telling you this is not how you should be doing this; you are not really solving the problem [of automated pricing intelligence].*
>
> - Top-level executive at PriceWise

Yet, they went ahead with forming, hiring and continually expanding their QA team throughout their 10-year journey as a company. The demands of clients to deliver high-accuracy results and the limits of what machine learning models could achieve forced them to make peace with the sub-optimal route.

> *Just because you have the motivation to solve a certain problem [of automated pricing intelligence], doesn't mean that certain technical problems are actual business problems. The expectations from the customers in terms of speed, in terms of accuracy keeps on increasing. We have to guarantee 95% accuracy. But the reality is that at this stage of tech maturity, none of the machine learning models are going to give you that kind of accuracy… which means humans still have to be involved… it has to be a human in the loop…*
>
> - Top-level executive at PriceWise

As engineers and technologists who set out to build cutting-edge pricing intelligence, PriceWise's leadership recognized AI and ML work as foundational and high-valued expertise within the company. Data labelling work done in the QA team was a practical necessity and seen as a technological compromise they could not avoid. This outlook was observable as much in how they described the role of the QA team during interviews as it was in the lack of appreciation for the QA team during the 10-year anniversary celebrations. Working in such an environment, QA analysts also saw their role as unavoidable "manual" work that was not as significant or valuable as "technical" roles in model development or data engineering. Consequently, they sought to move out of data labelling and into mainstream technology roles involving programming.

5. Discussion

Among AI researchers, developers and entrepreneurs, data work is generally undervalued in comparison to model development, even though they all recognise and acknowledge the significance of quality data in building AI systems (Sambasivan et al., 2021). Data labelling and annotation work represents this paradoxical relationship between data and model within AI production through specific configuration of human and the machine in data work. While the annotated and cleaned data are prerequisite to create sophisticated ML/AL models, the human labour that produces this clean data cannot be acknowledged fully if the techno-entrepreneurial vision of "self-learning" machine intelligence is to be celebrated and sustained. As a result, what we see is a creation of an underbelly of AI production conditions that thrives on human data workers on a mass level and yet masks them through different mechanisms of microwork, ghost work and remote outsourcing (Gray & Suri, 2019; Irani, 2015b). The paradox also lies in the fact that an artificial intelligence that functions through human labour can only claim its intelligence by hiding that labour (Ekbia & Nardi, 2017). The only human labour that gets acknowledged in AI production are those working on algorithms and models, as is evident not only in the literature (Sambasivan et al. 2021; Irani, 2015a) but also in our study, from the 10-year anniversary celebrations at PriceWise, where the teams that work on computer vision- and machine learning-models receive recognition and earn appreciation.

This paradox both shapes the everyday work practices of the data labelling team and also the perception of their work as “low-skilled” “manual” by both management and workers themselves. Contrary to this perception, while analysing everyday work practices of data workers, we see that human intelligence and labour is constantly needed not only to prepare the data for training the model but also to verify the model’s ability to process data correctly even after training. Our findings show that both matching products and verification of algorithmically matched products involve high levels of human attention and discretion to produce error-free pricing analytics. Moreover, within the QA tool, which is the main site for data labelling work, any improvement (such as providing more suitable suggested matches) in the tool is believed to be due to improvements in the model rather than a consequence of the data workers’ efforts who improved the model in the first place through their constant feedback in the pre-QA stage.

Thus, despite massive amounts of human work throughout the day, everything seems “automatically” done and “improved” in the tool. As the main interface connecting the work processes of the model development team to that of the data workers, the QA tool served to mask the input of the data workers for the development team, creating a perception of a seamlessly automated system. This perceived “automation” aligns perfectly well with the vision of optimal systems that should rely less on human intelligence. Hence, even when the founders believed their “human-in-the-loop” technique made their AI product (pricing intelligence) more accurate and thereby more attractive to clients, they regarded the approach as “sub-optimal”, reluctantly resigning to the limitations of machine learning models to generate high-quality output without any human intervention. Reliance on human labour as a “sub-optimal” alternative not only reflects a managerial devaluation of “non-model” work as mundane and menial but also produces structural conditions shared by the data workers themselves who hardly see their work being significant for model development and keep craving for meaningful “technical” roles in AI production. Much of their feelings of becoming redundant in the company gets further validated when engineering teams get applauded for their “high-value” work in computer vision.

Hence, we see that the techno-managerial vision of AI as “intelligent” and “automated” needs to mask its human in the loop technique. Further, framing data work as “low-skilled” and “sub-optimal”, almost as if they were workarounds that need to be ‘fixed’ in the future, feeds off of and lends continued relevance for that vision. In pursuit of a normative position of what AI should be, we are left with a denial of how AI is actually produced now.

6. Conclusion

In our critical enquiry of data work, we present the everyday work practices of data labelling and reflect on the underlying human-machine configurations. We also highlight how these configurations lend and sustain the perceptions of seamless functioning of AI, fulfilling the techno-entrepreneurial visions of “intelligent” systems. Our work adds to an emerging body of scholarship on data labelling that calls attention to two important aspects of this new-age technology work within the AI industry. First, we bolster the call to shift attention from the “future of work” discourse to a more nuanced conversation about current work conditions in the production and development of AI. We believe that as long as we continue to orient ourselves to the future implications of deployment of AI at workplaces (Posada, 2020), we will treat these implications as after effects (unintended at times) of an already finished and “irreversible” AI system and not as a problem embedded within its production that must be addressed at present. Such treatment will hardly challenge the way AI is produced and thereby question why those harmful implications for workers, such as, lack of autonomy, violation of privacy and increasing surveillance, occur in the first place. We show, for example, how the reduction in autonomy for data workers is managed through a tight orchestration of the supervisory control and tool mediated performance management. We argue that such human-machine configuration that retains higher control over “low-skilled” data workers are required to maintain the perception of a “high calibre” machine learning model aided only by “high-skilled” data scientists and engineers.

Second, through our work, we aim to expand the scope of ethical AI to include the labour conditions and work arrangements through which AI is produced. Most scholarly conversations concerning AI and ethics focuses on the impact of AI on bias, discrimination, individual freedom, accountability and so on (Jasanoff, 2017; Taylor, 2017; Ustek-Spilda et al., 2019). Even when ethical concerns spill over to AI development, they mostly look at biases in the data processing and model development, attributing those biases either to individual workers or to training datasets (Floridi, 2016). However, we contend that labour processes of AI development, including data labelling, content moderation and other similar work practices that mask, devalue and disepower human workers warrant scrutiny and inclusion in the ethical AI agenda. Recent work (Miceli et al., 2021; Sambasivan et al., 2021) in this vein have begun revealing how biases in datasets arise through imbrication of human practices that are mediated through the technological tools and apparatuses. We highlight how specific organisational structures with their normative vision favour certain kinds of human-machine configuration in data work. This plays an equally important role in shaping the conditions of data processing, which may likely have knock-on effects for biases or other issues of data quality. By engaging with work arrangements in AI production, we aim to confront the underlying conditions of its production that lead to ethical concerns and thereby push the boundaries of ethical AI discourse.

7. References

Adams-Prassl, J. (2020). When your Boss comes Home: Three Fault Lines for the Future of Work in the Age of Automation, AI, and COVID-19. *Ethics of AI in Context*, 1–11.

Ajunwa, I., & Greene, D. (2019). Platforms at work: Automated hiring platforms and other new intermediaries in the organization of work. In *Work and labor in the digital age*. Emerald Publishing Limited.

Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30.

Bilić, P. (2016). Search algorithms, hidden labour and information control. *Big Data & Society*, *3*(1), 2053951716652159. https://doi.org/10.1177/2053951716652159

Charmaz, K. (2017). The power of constructivist grounded theory for critical inquiry. *Qualitative Inquiry*, *23*(1), 34–45.

Crawford, K., & Joler, V. (2018). Anatomy of an AI System. *Retrieved September*, *18*, 2018.

Ekbia, H. R., & Nardi, B. A. (2017). *Heteromation, and other stories of computing and capitalism*. MIT Press.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160112. https://doi.org/10.1098/rsta.2016.0112

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280.

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*, *23*(2), 135–162. https://doi.org/10.1177/1024258916687250

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

Gupta, H. (2019). Testing the future: Gender and technocapitalism in start-up India. *Feminist Review*, *123*(1), 74-88.

Irani, L. (2015a). The cultural work of microwork. *New Media & Society*, *17*(5), 720–739.

Irani, L. (2015b, January 15). Justice for "Data Janitors." *Public Books*. https://www.publicbooks.org/justice-for-data-janitors/

Irani, L. (2015c). Difference and dependence among digital workers: The case of Amazon Mechanical Turk. *South Atlantic Quarterly*, *114*(1), 225-234.

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society*, *4*(2), 2053951717724477.

Joshi, S. (2019, September 9). How artificial intelligence is creating jobs in India, not just stealing them | India News—Times of India. *The Times of India*. https://timesofindia.indiatimes.com/india/how-artificial-intelligence-is-creating-jobs-in-india-not-just-stealing-them/articleshow/71030863.cms

Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., ... & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, *7*, 2378023121999581.

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1603–1612.

Miceli, M., Schuessler, M., & Yang, T. (2020). Between subjectivity and imposition: Power dynamics in data annotation for computer vision. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–25.

Miceli, M., Posada, J., & Yang, T. (2021). Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? *ArXiv Preprint ArXiv:2109.08131*.

Miceli, M., & Posada, J. (2022). The Data-Production Dispositif. *arXiv preprint arXiv:2205.11963*.

Moore, P., & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *New Media & Society*, *18*(11), 2774–2792.

Murali. (2019, March 21). How India's data labellers are powering the global AI race. *FactorDaily*. https://archive.factordaily.com/indian-data-labellers-powering-the-global-ai-race/

Murgia, M. (2019, August 8). AI's New Workforce: The Data-Labelling Industry Spreads Globally. *Financial Times*. https://medium.com/financial-times/ais-new-workforce-the-data-labelling-industry-spreads-globally-f472cb1bac09
Pietrobelli, C., & Rabellotti, R. (2011). Global value chains meet innovation systems: Are there learning opportunities for developing countries? *World Development*, *39*(7), 1261–1269.
Posada, J. (2020). The Future of Work Is Here: Toward a Comprehensive Approach to Artificial Intelligence and Labour. *ArXiv Preprint ArXiv:2007.05843*.
Poster, W. R., Crain, M., & Cherry, M. A. (2016). 1. Introduction: Conceptualizing Invisible Labor. In *Invisible Labor* (pp. 3–27). University of California Press.
Raval, N. (2021). Interrupting invisibility in a global world | ACM Interactions. *Interactions*, *28*(4), 27–31.
Roberts, S. T. (2016). Commercial content moderation: Digital laborers' dirty work. In Noble, S.U. and Tynes, B. (Eds.), The intersectional internet: Race, sex, class and culture online (pp. 147-159). New York: Peter Lang.
Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15.
Shestakofsky, B. (2017). Working algorithms: Software automation and the future of work. *Work and Occupations*, *44*(4), 376-423.
Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2), 2053951717736335.
Ustek-Spilda, F., Powell, A., & Nemorin, S. (2019). Engaging with ethics in Internet of Things: Imaginaries in the social milieu of technology developers. *Big Data & Society*, *6*(2), 2053951719879468. https://doi.org/10.1177/2053951719879468
Wood, A. J., Lehdonvirta, V., & Graham, M. (2018). Workers of the Internet unite? Online freelancer organisation among remote gig economy workers in six Asian and African countries. *New Technology, Work and Employment*, *33*(2), 95–112.
Woodcock, J., & Graham, M. (2020). *The gig economy: A critical introduction*. Polity.