Tracing the Displacement of Data Work in AI:
A Political Economy of "Human-in-the-Loop"

BIDISHA CHAUDHURI
UNIVERSITY OF AMSTERDAM
THE NETHERLANDS

SRRAVYA CHANDHIRAMOWULI
UNIVERSITY OF EDINBURGH
UNITED KINGDOM

Abstract

In this study, we trace the evolution of a data work team in an artificial intelligence (AI) startup in India. By bringing attention to data work, which is the indispensable work of preparing annotated datasets for training AI systems, conducted within a formal organisational set up, we underline: 1. how organisational approaches adopted to balance investor and client preferences shape work arrangements and spatial division of the data workers; 2. how relations between the data team and the 'core' technical team serve to invisibilise human labour in the production of AI; and 3. how increasing codification of data work leads to devaluation of data work within the organisation and deskilling of young data workers at large, making them vulnerable in choosing a meaningful career path of their choice. In tracing this trajectory of displacement of data workers employed in a formal sector, we show that the prevalent characterisation of data work as being invisible or precarious is not inherent to AI nor inevitable in its labour processes. Rather, it is produced through the specific embedding of AI production within the political economy of startup capitalism. Through this, we seek to recentre the discourse on AI and future-of-work away from deterministic projections of AI's impact on work and towards the specific labour processes of AI and its implications for the skills and career trajectories of a young and growing workforce in the Global South.

Keywords

Introduction: Artificial Intelligence (AI) and Human Work

The burgeoning discourse on AI and the future-of-work offers two drastically opposing yet equally deterministic views about the impact of AI on human labour (Shestakofsky 2020). On the one hand, there is the widely held belief that AI will alter the nature of work and hence the need for human skills (or at least certain types of human skill) associated with it (Autor, Levy, and Murnane 2003). On the other hand, automation anxiety (Akst 2013) fears job-loss, deskilling, and the replacement of human

jobs by new-age AI. However, as Munn ([2022](#)) points out, automation is nothing but a mythical rhetoric rehearsing a centuries-old fantasy of rendering human workers obsolete, where AI is merely the latest addition to that rhetoric. Despite its fictional overture, this rhetoric warrants critical scholarly attention not only for its influence on public discourse on automation, but for its ability to invisibilise human workers and the structural conditions of their work and shift our focus to a "fantastical" future of total and complete automation. In this paper, we concentrate our analytic gaze to a set of human workers who form the building blocks of current AI models — the data labelers (or data workers in general), and the political economy of the labour process within which data work takes place. In unpacking the labour processes of data work, our study bolsters the calls ([Posada 2021](#)) to shift attention from the rhetoric of the future-of-work to the current organisations of AI and (human) work. Through this, we aim to contribute to the AI and labour discourse in two ways: first, we bring attention to a human-labour intensive component of AI work, which is often neglected; and second, by analysing the labour process that organise data work ([Ilavarasan 2008](#)), we reveal how this neglect is structurally produced rather than being an inherent feature of the technology. By examining these labour processes, we find that data work becomes gradually displaced within the AI ecosystem.

We show that this displacement happens through two simultaneous and intersecting labour processes. First, the spatial displacement of data work through platformisation[1] or outsourcing ([Gray and Suri 2019](#); [Miceli and Posada 2022](#)) as well as organisational and sectoral hierarchies that separate data work from model work[2] ([Irani 2015a](#); [Sambasivan et al. 2021](#)). Second, the social displacement[3] of data workers by devaluing their skills as mundane and pushing their location into the bottom of the hierarchy of AI-workers' skill trajectories. This contributes to the social construction of data work as not being skilled, shaping how data work and those performing it are perceived, thereby socially constructing the perception of skills involved in data work and of those performing those data work.

While both spatial and social displacement of (data) work is well documented in the literature (as we will show in the following section), through our study of data work in an AI startup in India, we empirically elucidate that these two processes of displacement intersect and influence each other. Through the detailed observation of everyday data work and its position within the

---

[1] By platformisation we mean gig work done through platforms such as Amazon Mechanical Turk and other similar platforms.
[2] This separation between data and model work is akin to Taylorist separation between planning and doing and which continues to thrive as Digital Taylorism emphasises" on vertical chains of information and control devalues horizontal interactions between those chains which otherwise might allow staff to learn directly from their peers" ([Cole, Radice, and Umney 2021, 91](#)).
[3] By social displacement we mean downgrading a group's social status in terms of prestige and power within a bounded social space ([Chernoff [2010] 2013](#)).

organisational labour processes, we show the specific ways in which these processes of displacement affect those involved in the data work. By expanding on the spatial and social concepts of displacement and their implications for skill development, we aim to move beyond the tropes of 'invisibility' and 'precarity' through which data work is characterised and focus on the structural scaffolding of AI production, particularly in the Global South.

The article is divided into three sections: the primary research section draws on insights from the literature that highlights the process of displacement of human workers within the larger discourse of automation and within AI-based automation in recent times; the methodology section describes our ethnographic fieldwork among data workers in an Indian AI start-up and our data analysis methods; and then we present the labour processes within which data work is performed and its implications for skill trajectories of data workers. Following that, we discuss how these specific labour processes and skill trajectories connect to the broader political economy of AI work in India. Amidst increasing demand for data work and its celebrated status in the Global South for creating livelihood opportunities (Joshi 2019; Murgia 2019), we aim to highlight the nuanced and substantive changes in the labour processes that produce AI and algorithmic technologies and its impact on skill transformation within the AI industry that affects the long-term career prospects of human data workers.

Prior Research
Structural Organisation of Human Data Work in the Production of AI

Most advanced algorithmic systems designed to make predictions and recommendations or to automate tasks are fuelled by vast datasets. These datasets must be cleaned, labelled, and verified — tasks that depend on human effort (Bilić 2016; Irani and Silberman 2013; Tubaro, Casilli, and Coville 2020; Chandhiramowuli and Chaudhuri 2023). However, the language of 'cutting-edge innovation', 'smart' technologies and 'data-driven' or 'intelligent' solutions masks the human labour necessary both to prepare data and to configure the machines and models that use them (Gray and Suri 2019; Irani 2015a; Newlands 2021). Job roles such as those offered by the Amazon® Mechanical Turk platform, content moderators, and data workers represent the new age of the information services economy (Gray and Suri 2019). This human (data) work, which is foundational to the AI industry, takes place in gigs on platforms as well as through outsourcing from the Global North to varied geographical locations in the Global South (Miceli, Schuessler, and Yang 2020; Natarajan et al. 2021; Mehrotra 2022; Wang, Prabhat, and Sambasivan 2022), resembling the IT services offshoring, that began since the early 2000s (Upadhya [2011] 2020; Upadhya and Vasavi 2008).

Gray and Suri (2019) observe a phenomenon that they call 'the paradox of automation's last mile', wherein AI-based automated systems envisioned and developed to replace human workers often tend to create new tasks for humans. As AI is adopted into newer domains, the ever-moving target of 'full automation' continues to depend on human labour (Crain, Poster and Cherry 2016). We see this, for instance, in Poster's (2016) study of virtual receptionists, which shows how advanced menial automation strives to replace human workers, first, with the use of digitally-manufactured

virtual assistants to (in)visibilise selectively (or strategically as seen in Newlands (2021)), human-like characters of sociability, emotion and interactiveness, and second, by leveraging communication network technologies, to disperse human receptionists spatially. Both these developments happen in parallel across different locations within the global information economy.

This paradox of the invisible yet indispensable role of human labour in the current phase of AI-based automation, unfolds in specific ways in the Global South. Far from being invisible, the so-called 'behind the scenes' (Roberts 2016) jobs in data work are not only highly visible but also sought after in the Global South as they represent dignified service sector livelihood opportunities for large sections of its youth (Raval 2021). This was true for BPO (Business Process Outsourcing) workers in the early 2000s and is true for the current wave of data workers. However, the data workers of the AI industry in the Global South, much like their predecessors in the outsourcing industry, are located at the lower end of the global value chain of AI production, separated from the 'core' processes of knowledge production (Pietrobelli and Rabellotti 2011; Graham, Hjorth, and Lehdonvirta 2017; Irani 2015b). The fast-growing AI industry in the Global North employs thousands of young workers from the Global South to form the backbone of the AI ecosystem (Murali 2019), while distinguishing and distancing them from the modelling work of AI (Sambasivan et al. 2021).

This displaced work is typically cast as 'mundane', 'repetitive', and 'non-cognitive' and performed by workers overqualified for the job, yet turning to it in the face of rising unemployment (Graham, Hjorth, and Lehdonvirta 2017), leading to the devaluation of the skills of those who perform data work (Shestakofsky 2017, 2020). Thus, data workers become displaced both by moving behind a screen (Roberts 2016) or geographically far away from the technical core (Irani 2015b) either through outsourcing to different locations or through creating organisational structures that separate the modelling work from the data work, and by moving down in the skill hierarchy, the new division of labour between humans and the machine within the discourse of automation (Ekbia and Nardi 2014, 2017). This displacement draws on and feeds into the perception that modelling work is more desirable and highly esteemed than data work in AI (Sambasivan et al. 2021).

Skill Question in Data Work

However, analysing skill trajectories is a tricky endeavour, since, far from being objective, 'skill' is a contentious category (Baum 2008; Rigby and Sanchis 2006). As social constructs, skills are produced through social processes that reflect the dominant political economy and existing structures of power (Baum 2008; Rigby and Sanchis 2006). For example, feminist scholars show how women have been historically excluded from job categories labelled 'skilled' (Hicks 2017; Warhurst, Tilly, and Gatta 2017). Thus, the value of skills is not just measured through a worker's technical capacities and professional expertise but also using political economic factors, such as: which jobs are being created through new technologies; where does the capital come from; how does the capital delegate work between technology and human labour; who performs those jobs and where, and what kind of training is required (Aneesh 2001; Baum 2008; Shestakofsky 2020). Consequently, what constitutes

deskilling and reskilling[4] becomes contested, resulting in multiple conceptualisations such as the 'degradation of work' (Braverman 1974), the separation or loss of skill from degradation of work (Cockburn 1999), and 'skill saturation' (Aneesh 2001).

Braverman (1974) attributes the overall degradation of human work within industrial capitalism to the detailed division of labour and associated codification of work processes. While he does not use the term deskilling, he hints at overall reduction in autonomy and control of human labour through codification of work. Aneesh (2001) reconceptualises this transformation in the post-industrial economy with regard to increasing use of information communications technology in service sector employment. He encounters a process of 'skill saturation' in service jobs, where human skills are 'exhaustively ordered' to generate predictable work processes and outcomes without much room for human creativity. We see a similar trajectory of codification and skill saturation in data work, made possible by limiting the scope for human/worker subjectivities in matching and verification that are feared to introduce biases in the dataset used for training AI systems (Hube, Fetahu, and Gadiraju 2019).

The first set of literature captures different ways in which data work is organised across social, geographical and organisational hierarchies and the second set of literature focuses on how these different forms of social and spatial organisations affect the notion of skilling among data workers. We draw on both these strands of literature to explore how data work is positioned within the Indian AI industry and its global linkages; how data workers relate to other AI workers; and how the everyday performance of data work is organised and controlled (Ilavarasan 2008). We address these questions within the context of the budding data annotation industry in India, through careful scrutiny of data work in an AI product start up in India's Silicon Valley, Bangalore. Addressing these questions about human data work allows us to identify the specific mechanism of spatial and social organisation of data work and its implications for the human workers involved in this work within the specific context of AI production.

Methodology

Field Site: An AI Startup in India

We refer to our field site as *PriceWise*, a Bangalore-based company that was founded in 2011 to develop pricing analytics for e-commerce brands and retailers across the world that seek to identify the best prices and promotions they can offer to attract consumers and grow their businesses. *PriceWise* is in

---

[4] Drawing on Braverman (1974), we refer to deskilling as a process of subordination of workers' skills and knowledge through increasing mechanisation of routine tasks achieved through technological innovation in the production process. Contrary to this theory of deskilling, a more optimistic take on technological innovation focuses on the notion of reskilling which refers to the process of learning new skills, often with the intent of moving into new roles or occupations (Lewis 2021).

the business of informing these pricing decisions through AI-driven data analytics; its clients are e-commerce retailers and consumer brands operating primarily in North America but also in India and other large economies. The company uses AI to provide data insights (or pricing intelligence, as it is called) that inform competitive pricing decisions, identify emerging market trends, and develop promotion strategies.

To arrive at this 'intelligence', a key underlying step is product matching — identifying the exact same products across client and competitor platforms, so as to compare their pricing. *PriceWise* uses AI techniques, especially computer vision and machine learning, in its product matching algorithms to process thousands of its clients' products and identify relevant matches for price comparison. An integral part of developing this AI-based product matching system is training it with annotated datasets.

PriceWise embraces the 'human-in-the-loop' approach to building AI-driven pricing analytics, where human data workers play a crucial role in creating training datasets for AI models, verifying their results, and completing those tasks that were not satisfactorily done by AI. Their efforts are crucial to achieve the promise of 95 per cent accuracy that *PriceWise* offers its customers for its pricing intelligence.

PriceWise created the Quality Assurance (QA) team to work on these tasks. They typically recruit recent graduates, primarily from engineering and other STEM backgrounds (but in recent years also from commerce and humanities) and train them in finding and verifying product matches. At the time of our study, the QA team comprised of about 20 analysts[5] across four sub-teams, each managed by a Team Lead (TL) and a Backup Team Lead (B-TL). TLs are responsible for allocating work, maintaining delivery timelines, and ensuring the accuracy of day-to-day work. TLs are typically senior QA analysts with over 4–5 years of experience. The B-TL takes over for the TL in their absence but is otherwise tasked with mentoring new team members and carrying out data verification. The QA team conducts their everyday work of product matching and verification through a web-based tool, simply referred to as the QA tool. This tool is the main site of data annotation work where datasets would be made available and then assigned to the analysts by TLs, who also monitored the progress of work using the same tool. Thus, the QA tool is central to the everyday work practices of the QA team.

Research Design
We deployed a constructivist grounded theory approach ([Charmaz 2017](#)) to examine emergent job roles in data work and their associated work practices in an exploratory yet reflexive manner. We

---

[5] The members of the QA team held a few different designations indicated in this section and they were typically referred to as analysts. In our writing, we use the terms QA analyst and data worker (and similarly QA work and data work) interchangeably.

conducted interview- and observation-based ethnographic fieldwork at the in-house data labelling team of *PriceWise*.

We came in contact with *PriceWise* through institutional networks from the institution where this research was based. We held a couple of preliminary discussions with the senior leadership team at *PriceWise* to present our research interests in understanding the role of human labour in AI-driven automation and negotiate access to the QA team. We began our study by conducting an open-ended, unstructured focus group discussion (FGD) with *PriceWise's* top-level management; this discussion shed light on the organisation's history, pivotal shifts in the company's trajectory, and their approach to data work, perceptions of AI, automation, and the human-in-the-loop approach. We also conducted an FGD with the manager and the TLs of the QA team to gain an initial understanding of the work arrangements of data work within *PriceWise*.

Following these initial discussions, the second author spent 6 weeks[6] between March-to-May 2021 as an intern in the QA team, working alongside data workers at *PriceWise*. This participant-observation method allowed us to engage actively in everyday data work, access the tools they used, observe team dynamics and work routines of the QA analysts, and interact with QA analysts with varied levels of experience and seniority within the organisation. In particular, being embedded within the team as a QA analyst provided the opportunity to gain hands-on experience of the QA tool as well as learn about its evolution through the experiences of other members of the QA team. Due to the Covid-19 pandemic, the company worked remotely at the time of the study. This meant that our interactions with *PriceWise*, including interviews and participant observation, were primarily conducted online.

We conducted 20 in-depth interviews to complement the observations. We spoke with QA analysts (9), TLs (3), B-TLs (4), the QA manager, and former QA analysts (3) who moved into other teams within *PriceWise*. Of the 20 interviews, eight were with women in the team, including three senior members — approximately representing the gender-ratio in the team at the time of our study. In the interviews, we explored the roles and responsibilities of workers within the QA team, their educational background, career trajectories and professional aspirations, and the tools used as part of QA processes. During interviews, we clearly described our professional and educational background, prior experience, and purpose of the study to clarify our position as external researchers without any affiliation to *PriceWise*. We also obtained informed consent orally and have pseudonymised the names and other identifiable details of the people, products, and the company to ensure confidentiality.

---

[6] This research access to conduct ethnographic fieldwork was facilitated through a non-disclosure agreement (NDA) signed between *PriceWise* (a pseudonym) and our institution to allow access to the researcher to be embedded in the QA team for research purposes.

Labour Processes of Human-in-the-loop in Silicon Valley of India: A Start-Up Story

In this section we illustrate how data work is positioned within an AI startup, *PriceWise*. We take a longitudinal view of data work at *PriceWise*, bringing attention to the ways in which data work evolved over the course of a decade. In particular, we highlight a) the shifting position of data work within an AI startup and the role of investors in shaping this, b) the consequences this bore for the organisational mobility of data workers and finally, c) the broader implications for the skills and career trajectories of data workers in the long run.

'Forcing Function' of Business and Start-Up Capitalism: Growth of *PriceWise*

When *pricewise* started out a decade ago in the early 2010s, their rudimentary AI-based algorithms for product matching—upon which pricing analytics was contingent—required significant human involvement in data annotation. This led to the creation of the QA team, whose analysts verify the algorithm's product matches and manually find matches for those products for which the algorithms failed to find a match (we elaborate on these tasks in more detail in a later section). In their initial days, when the employee count at *PriceWise* was just around a dozen, about five-to-six of them were QA analysts engaged in the product matching and verification work.

> Initially, there was a fair bit of manual effort required. For instance, if it [the product matching algorithm] was at 50 per cent [accuracy], the remaining 50 per cent has to be identified manually, cleaned, etc. We realised within the first couple of years that we need people for this. (Ramesh, part of the senior leadership team)

Over the next couple of years, *PriceWise* gradually expanded their development teams, and improved the accuracy of their product matching system for the initial product categories and websites they processed. But these improvements in algorithmic performance happened alongside the expansion of its business as well. The company's client base grew to include more brands and retailers seeking pricing analytics against their competitors. With that, the range of brands, products and websites (including those in languages other than English) that the product match algorithms had to process increased significantly. Further, alongside increasing its client base, *PriceWise* also diversified the analytics it offered clients to include branding and promotions. This introduced newer products as well as newer parameters to account for in the product matching algorithms. The goalposts for the performance of the product matching algorithms thus kept shifting, placing practical limits on achieving and maintaining high accuracy in the matching results, and necessitating a continued reliance on the QA team to fill the gaps in algorithmic performance.

The QA team conducted manual verification and product matching work, providing annotated training datasets necessary for improving the models, while simultaneously ensuring high accuracy results for clients. They thus became an integral part of *PriceWise's* workflow and grew in size as the company's business expanded. After four-to-five years, by 2015, it became the biggest

team in the company, with 30–40 members.[7] This 'human-in-the-loop' approach played a crucial role in *PriceWise* gaining an edge over competitors who relied on 'pure' AI-driven systems. The Manager of the QA team, Samantha, was confident that the team was in no threat of being 'automated away'. Far from becoming redundant, the volume of data work her team handled had only increased as the algorithms improved, and the business grew.

> I really would not say that anytime down the line, we can remove humans from this loop. Because the system cannot be constant, it has to be dynamic and improve *continuously*. The input for how to improve the system has to come from humans [QA analysts]. The world is changing so much, we are getting more categories online, so we are constantly training or verifying the systems on these new categories and providing feedback to improve. Recently, we've also been getting into other domains like travel, and hospitality. While getting into such new things, it again requires humans to improve the system. (Samantha, Manager, QA team (emphasis added))

Samantha further clarified her point by citing how as the company grew, the reliance on the QA team also expanded. However, this continued reliance on data work was not devoid of tensions — during our study, we found that though the company's growth necessitated a continued reliance on the QA team, it also brought significant changes to *how* the growing QA team and its members were positioned within the company. When *pricewise* onboarded their first large e-commerce client, it increased QA workload significantly, leading to the creation of a night-shift QA team, staffed by temporary, contractual workers, to handle the additional workload and support the main QA team. Once the major deliverables for this large and important client were met, *pricewise* dissolved the team and absorbed only a few 'bright' workers into the regular QA team. By 2016–17, *PriceWise* had secured a major round of funding and also set up offices in the US to cater to the North American client base. One of our interlocutors mentioned that it was around this time that *PriceWise* looked to increasingly outsource QA work to third-party vendors in tier-2 cities and shrink their internal QA team.

Besides outsourcing projects and downsizing the internal QA team, *PriceWise* also stopped employing QA analysts directly and instead hired them as contractors through a consultancy service.[8]

---

[7] We learnt that the QA team had been the one of the largest teams in terms of team-size through our interview with a former QA analyst who had moved to another team within *PriceWise* at the time of our interaction with him. We corroborated this in subsequent interactions with senior members of the QA team. However, we did not have access to exact data (either through our field work or publicly available online resources) on team-wise composition and total employee count through the years of *PriceWise's* expansion.

[8] A consultancy service provided required human resource on-demand to companies like *PriceWise*. They employ and manage the workers and assign them to clients like *PriceWise*. In this route, the workers join the client company's team as a vendor and work alongside the team internally, while being paid and managed by the consultancy. The consultancy workers typically work for one client at a time and for extended periods (at least a few months) and strive to be absorbed by them as their employee due to the

Under this arrangement, the internal QA team was a mix of full-time *PriceWise* employees and contractual workers recruited from a consultancy. For the company's top executives, these shifts were almost inevitable:

> As we grew, obviously *the business itself puts that forcing function* where you gotta scale! But as the business grows, we should not be growing the people linear, that has to be sublinear, especially for QA. That's the *mandate for us*! We need to consider things like revenue per employee, optimise for cost and there is a margin we have to look at … (Ramesh, part of the senior leadership team (emphasis added))

In describing the organisational shifts in QA as a 'forcing function' of the business, it was implied that they were an inevitable consequence of a business growing and scaling. The 'mandate' that underlies these shifts came from the company's investors. As a startup, *PriceWise* relied on venture capital (VC) funding, raising two rounds of funding over ten years. The involvement of investors introduced organisational metrics and business targets, and scrutinised *PriceWise's* technological choices and approaches, thereby influencing the priorities and goals of the company.

> If you are looking for investors, they care a lot more about this [use of AI or automation]. Investors care about things like: are you using the latest technology or not? How are you going to scale and get a lot more customers? (Veer, part of the senior leadership team)

We see the logics and pressures of funding operate at two levels: First, the investor-driven imperative to use the latest technology compels companies to lean heavily into the vocabulary of automation, promising to deliver intelligent machines that can accurately model an ever-changing world. Data workers, like *PriceWise's* QA analysts, are then enlisted in the service of AI (or automation) to fulfil these ambitious promises. Second, as *PriceWise* fulfilled and delivered the semblance of automation and as a result, expanded and acquired more funding to build their AI business further, the logics of venture-capital funding placed pressures to displace their QA team. Thus, we see the financial imperatives, particularly the funding model of *PriceWise* as a startup reliant on investor funding, bearing a significant influence on both the continued reliance on data work as well as their organisational displacement.

By tracing the scaling up of *PriceWise* through the years and the evolution of the QA team along with it, we see that there was nothing inevitable about the AI technologies that led *PriceWise* to organise the data annotation within the firm in the way that it did. It was shaped by their quest to project themselves as an AI product company and attract VC funding as such. In our conversations with the leadership team of *PriceWise*, their ambition to be a product company (and not an 'Indian' IT

---

instability of being a contractual worker and limited benefits of being employed with a consultancy service.

service company) that could tap into the Silicon Valley funding ecosystem was evident. This entrepreneurial ambition and their subsequent decision to raise funding fits well within the start-up ethos and ecosystem in Bangalore, which has 'begun to self-consciously transition from a site of 'back-end' digital labour to fashion a new affective and material future for itself as a 'Start-Up City'' (Gupta 2019, 76) with much support from the state to create a conducive environment for innovation and entrepreneurship (Gupta 2019). Similar to Shestakofsky's study based in California (2017, 2020), we find the start-up ecosystem of Bangalore and associated funding cycles shaped the trajectory of *PriceWise* and the specific labour processes it introduced to organise data work. In the next section, we illustrate how the members of the QA team experienced these shifts in the organisational position of their team.

Sitting Side-by-Side: Organisational Position and Mobility of Data Workers

The QA team at *PriceWise*, at the time of our study in 2021, comprised about two dozen people in their early twenties; for many it was their first job. They typically held a degree in a STEM field, although this was not a hard requirement. Most of them aspired to (eventually) move into web development and data engineering roles within *PriceWise*. It was this possibility to transition to these other roles that brought many of the QA analysts to *PriceWise*. The company not only recognised this aspiration but also sought to leverage it. Stories of how former QA analysts had moved from doing 'manual testing' — like work in QA to roles involving more 'technical'[9] work were often shared to highlight how the company supported its 'humans in the loop'.

Kishore was one such employee. He joined *PriceWise* as a QA analyst, an early employee of the company and the second person recruited to the QA team. After two years in QA, he moved to the data engineering team, into roles that required scripting and programming skills. A year later, he became a senior data engineer, eventually moving into the product development team. His successful transition from data labelling to 'technical' teams was an oft-repeated success story, cited as an example of the support the company provided its 'hard working' employees. Kishore, too, echoed the sentiment.

> I was working 16–17 hours in those days. I used to do QA work in half a day and the rest of the day *I used to sit with developers* and understand how they used to code, etc. If you are interested in coding, [you can] just talk to the developers, take some small tasks … I am pretty sure that the company will support you. *They will surely give* [*opportunities*] *to everyone*, as they have given me. The only thing is, you have to put in the effort from your end. That's it! (Kishore, former QA analyst (emphasis added))

---

[9] We indicate manual and technical within quotes to reflect that these are emic categories. As researchers, we are critical of these labels but retain them to reflect the values that our participants attached to them.

Kishore was referring to a time when the entire office had no more than 30 employees, all of whom worked together in one room. As QA analysts sat and worked alongside data engineers and developers, they got to know the different types of work within the organisation, and forge relationships with and learn from their co-workers. These networks and interactions facilitated their lateral movement across teams. However, as the company grew in size from a couple of dozen people to over a couple of hundred, formal team structures began to emerge and the distance between teams increased. QA analysts no longer sat alongside developers and data engineers; nor could they observe their work. If the QA work required interacting with software development or data engineering teams, it was only the manager or TLs who would be in contact with them. While this insight about the spatial rearrangement of data work within *PriceWise* was shared both by former and current QA analysts, the second author working as an intern analyst in the QA team also experienced this separation very clearly in the analysts' everyday work.

In the emphasis on the individual's initiative and hard work, as recounted by Kishore, and others like him, what gets obscured is how *PriceWise* as an organisation had scaled over the years as we discussed in the section above, and how that ushered in a clear separation between the QA team and the rest of the 'technical' teams. While the opportunities to transition to 'technical' teams still remained open, QA analysts no longer enjoyed the same levels of access to learn about these teams and their work as their predecessors earlier did.

This reduced opportunity for internal mobility of QA analysts was further constrained by the increasing contractualisation of QA work, described in the section above. When the contractors are brought onboard the QA team, they are told that they could potentially become full-time employees, in about 6 months, contingent on good, consistent performance. However, at least four contractual analysts told us that they had been working in the team for over a year but were yet to be made full-timers. The pandemic has been blamed for the delay. Effectively, it was, at the time of this study, nearly impossible for QA analysts to change teams within three years of joining the QA team. Their position within the company was delicately balanced on weak contractual obligations.

> Ultimately it also comes down to the number of seats we have in the office; for now, that is sorted. If the tech team starts growing, I may have to consider the vendor model [to reduce numbers in the QA team further]. There is multiple dynamics [at play] here. (Veer, part of the senior leadership team)

As QA work became more contractor- and vendor-dependent, it erased lateral movement opportunities for most of the data workers working in *PriceWise*. The logics of investment that dictate how QA work was accommodated (or not) in the company do not account for these lost opportunities for organisational mobility. It resulted in formal boundaries between teams that earlier allowed for more fluidity. As the QA team went from sitting side-by-side with the tech team to giving up their seats for them, we see a clear separation of 'innovative' work from 'menial' work (Irani 2015a; Sambasivan et al. 2021). Institutionalising this separation was a part of *PriceWise's* efforts to establish itself as a company of cutting-edge innovation in AI. In doing so, it reproduces in its own backyard

the very same hierarchy between product innovation and back-office support that it seeks to subvert. The fact that they relied on the human data workers and yet pushed them to the background had implications for the skills and career trajectories of those workers, which we explore in the following section.

Conditions of Everyday Data Work: Codification and Deskilling of Data Workers

The reduced opportunities for lateral movement created major concerns for data workers regarding their skill trajectories and career opportunities. Many of them held graduate degrees in engineering or other STEM disciplines. But their work, though squarely located within the IT industry, largely relies on their ability to search the internet, learn product categories on e-commerce platforms, and make sense of consumer products that were often inaccessible to them. And in this work, they use a proprietary software tool developed internally by one of the technical teams to streamline and codify their work practices as well as outcomes. While *PriceWise* did not necessarily restrict or prefer candidates with a STEM background, Samantha, the QA manager, revealed that the majority of the applications they received were from STEM graduates, which she believed was due to the company's status as an AI product company. When we asked some of the QA team members who had a STEM degree, they also alluded to the same sentiment. They considered the QA team as a steppingstone to more mainstream and lucrative roles and opportunities in the data science or AI job market in India.

Analysts are expected to login to the tool and get started with the day's work by no later than 10 AM. Upon logging in to the tool, they start working their way through the day's assigned list of products. One type of task for the QA analysts is to find the exact match, if available, for each assigned product from a list of close matches found by the matching algorithm. The QA analyst sifts through the algorithmic suggestions listed in the tool to find the exact product match for the base product. With each client, the kind of products and what counted as a match varied. On one occasion, a teammate alerted the second author to a wrong match they had added, pointing out that two otherwise identical skirts had different colour buttons! Moreover, products that are exactly the same could appear under different names on each platform. With experience, QA analysts became adept at identifying such similar-looking differences and different-looking similarities.

The tool is designed to curate the close, likely matches from the algorithmic output and present them for QA analysts to weigh in with their human discretion and determine the right match(es). Streamlining this over time facilitates training the product matching models which upon improvement, create a second type of task — data verification. Here, QA analysts verify matches made by the algorithm. During our interviews with the long-term members of the QA team, we learnt that the tool had been significantly expanded to align with increasing product verification work over

product matching.[10] As they carried out matching work using the tool, it provided data to identify and build new features into the tool that would eventually enable the shift from matching to verification work.

The tool is an important intermediary in the human-machine configuration of QA work, and it serves a crucial role in codifying and standardising the data work required for improving the models used by the engineering team and also in closely monitoring the day-to-day work and performances of data workers. It allows the company to standardise the collection of input from different kinds of workers — from inhouse QA analysts to contractual and outsourced workers. Further, the more the data annotators use this tool, the more they contribute to the future automation of their own work. Many of the workers we spoke to were already aware of this issue.

> If you see now we find matches when AI cannot. Apart from that, we also verify the AI output. But in future, when it improves, what will be left is just the verification part. Just reviewing the matches of the AI. (Chetan, QA analyst)

However, this possibility of a future automation is not their main concern. Their primary concern is the limited opportunity in QA to develop transferable 'technical' skills and the increasing gap between their skill and capabilities in the role of a QA analyst. As they primarily work on the tool that is proprietary to *PriceWise*, the skills they acquire are not transferable to other contexts. In our interviews, no analyst expressed any inclination to build a career in the data work domain. The lack of transferable skills and susceptibility to redundancy, motivated analysts to look beyond data work for professional development and as many of them held an engineering education background, they strongly aspired for that transition. Suresh, a QA analyst, was learning the programming language Python (on his own time), because he had heard that it was a valuable skill to find employment in the IT sector.

> We should have some technical work as well. The crawling done to curate product lists in the tool, for instance, why can't we do that crawling in the QA team itself? We can learn to do it, it will be useful for us also [as a skill] . . . Just doing 'tool' work is not useful for us. (Jeeva, QA analyst)

When they sought opportunities in other companies, they faced the challenge of demonstrating that they had acquired marketable skills and expertise in their role as a QA analyst.

> So many people have asked me, "you are a B-Tech [degree] holder, this [QA] seems to be a non-technical field, why are you not going to the technical side?" It's not easy to find openings for

---

[10] The role of the tool in this shift was facilitating and streamlining the annotation of product matches, which further improved the training data and enabled not only improving the performance of the models but also changed the nature of work of the QA team — from matching to verification.

someone like me who graduated earlier, has worked for 2 years but doesn't have technical experience. (Hansika, QA analyst)

Faced with these challenges, they view staying in *PriceWise* as the more viable choice and strive to leverage opportunities for lateral movement. This, many of them believe, is their only real shot at building a stable, long-term career in the industry. The position of data work within the lower end of organisational hierarchy and simultaneous confinement of data workers into rigid team structures and shedding of annotation burden through consultancy and outsourcing pose a severe challenge for their career aspirations. They realise the longer they stay in the QA team, their probability of getting stuck in this 'low-skilled' work not just in the company but in the industry at large, becomes higher as the gap between their qualification and the skill learnt in the job will keep increasing.

The conditions of everyday work of data workers at *PriceWise* elicit how the tool-based codification of data work leads to devaluation of data work as back-office services and thereby de-skilling those involved in the work by limiting their work to a specific proprietary software tool. In the following section, drawing on this case, we reflect on the broader political economy of human-in-the-loop work in India and its implications for the future of data workers.

Discussion: Political Economy of Human-in-the-Loop Work in India

With the rise in AI-driven businesses, data work has become a growing sector in India, expanding into diverse domains and providing livelihood opportunities. However, this growth is not straightforward. Studies found the gig-model of data labelling was no longer sufficient to meet the data needs of companies which led to a proliferation of domestic and transnational outsourcing (Wang, Prabhat, and Sambasivan 2022; Natarajan et al. 2021); such gigs not only served the large-scale technology companies from the Global North, but also catered to smaller companies and start-ups located in the vicinity of Bangalore (Mehrotra 2022). A recent report mapping the scope of the data labelling industry in India (Natarajan et al. 2021) identifies four broad patterns of conducting data work that characterises the spectrum of data labelling businesses within India:

1. Crowdsourcing or platforms-as-a-service (PaaS)
2. Software tools for data labelling (software-as-a-service)
3. People-as-a-service (to conduct labelling work as third-party vendors)
4. A combination of these services.

In its decade-long journey of building AI-driven analytics, *PriceWise* traversed this spectrum of data labelling businesses based on their labelling needs. For example, in their initial years with fewer clients, they largely managed their data work needs with a small in-house team. As their business scaled up, they began relying on third-party vendors that offered people-as-a-service. With the inflow of VC funding, they also built a proprietary tool to standardise data labelling across the lean in-house team, ad hoc contractual data workers, and multiple third-party vendors. This shows that there was no single or inevitable way of organising data work within AI; rather, multiple political

economic factors such as finance capital, labour supply, availability of relevant skills, the 'innovation' ecosystem, and the role of the state shape the spectrum of labour processes of doing data work for AI.

Gupta (2019) in her study of low-paid female professionals in a tech start-up in Bangalore, shows how these women workers are 'enrolled in experiments to test and perfect new technologies' that not only enable growth and profit of the company, but also allow the company to contribute to innovation through their labour. She calls this labour-induced experimentation and innovation 'office-as-laboratory' (ibid., 85). *PriceWise's* position within the Indian AI industry and the global supply chain makes it a formal 'office-as-laboratory', where the 'human-in-the-loop' technique is leveraged not only for product accuracy and profit making, but also to produce new training datasets for future AI automation. In this sense, *PriceWise* becomes emblematic of new start-up capitalism in the Global South that recalibrates the transnational outsourcing ecosystems for a new form of digital labour that becomes both subject and object of global technological experimentation (ibid.; Murphy 2017). As outsourcing became the dominant mode of positioning data work within the AI industry, BPO companies built their workforce by recruiting college-educated graduates to train and sustain AI systems. Thus, we see that the evolution of *PriceWise* as a company as well as its arrangement of data work with respect to its image as a product company (as opposed to a service company), was shaped by multiple factors of 'state ideologies, funding infrastructures and cycles, global entrepreneurial discourses and a postcolonial drive to situate Bangalore at the centre rather than the periphery of innovation.' (Gupta 2019, 78).

The positioning of *PriceWise* as a product company in Bangalore then made it an aspirational workplace for many young, STEM graduates who made up the QA team. It is important to note that the profiles of data workers would vary across the spectrum of labour processes in data labelling that we discussed above. This way of attracting a young workforce to data work in the 'IT city' shows many similarities between the contemporary data workers and the outsourced call centre employees of the early 2000s. Just as the call centre jobs created aspirational new middle-class workers with a clear sense of upward mobility and training in new skills of computer-mediated work (Upadhya [2011] 2020; Upadhya and Vasavi 2008), the data workers are also hailed as the new-age labour force in the rapidly growing global AI economy. Understanding this continuity in the trajectory of digital labour in India allows us to recognise the similarities in the structural conditions of capital allocation and labour processes that impacted employment through the IT service work in India then and the new milieu of data work now (Upadhya [2011] 2020; Gupta 2019). In both cases, limited skill trajectories meant that both generations of IT workers became (and in the case of data work, are susceptible to becoming) easily displaced and devalued in the long run as technology advances. Effectively, this meant that the boom for data labelling services absorbed a large population of young skilled people looking for a career in AI and created a process of 'dispossession [of skill] without exploitation' (Sanyal and Bhattacharyya 2009) within formal employment, where in they lose their future livelihood opportunities while being employed in specific kind of work at present time. The data workers in *PriceWise* were engaged in 'dignified' IT work with fixed hours, and pay, albeit with minimal opportunities for upskilling and career mobility. This process widened the gap between skills

(developed at work) and capabilities (prior training and educational qualification) of the data workers, making them more vulnerable within the job market and susceptible to joining the surplus labour in the Indian informal economy, which has been typical feature of Indian labour market after liberalisation (ibid.).

At *PriceWise*, as the AI system became better at identifying matches, human input was reduced to a binary choice: yes, it's a correct match or no, it's not. Here, the conditions of producing AI leads to the de-skilling of data workers. First, fresh graduates (even with specialised training) are hired only on the basis of their existing language fluency, basic understanding of the domain to which data belong (e-commerce platform in our case), and general digital literacy, such as, search, scroll and pattern recognition. Second, new workers receive only minimal training which is tied to the specific proprietary tool that mediates (as well as surveils) their work. Hence, we witness data work being treated as 'mundane', 'low-skilled' or 'unskilled' work. This devaluation of data work persists irrespective of whether it is done as a gig or as part of formal, salaried employment.

Within formal organisations, devaluation of data work emerges not only from the codification of the work through software tools, but also from its displacement within the organisational hierarchy along with the growth trajectory of the company. In *PriceWise's* early days, when the company was small, the QA team, which was completely in-house, had opportunities to observe, interact, and mingle with the rest of the company. These connections were crucial for finding and leveraging opportunities for lateral movement within the company. However, as the company grew in size, the QA team became increasingly separated from the rest of the organisation, initially due to formal team structures, later due to the company's shift to an outsourcing- and contracting-led model for data labelling. This meant that their visibility and access to other teams became limited or greatly slowed, thus limiting their 'upskilling' opportunities.

The shifts in the organisational structure reveal two important points. First, instead of being an (unintended) consequence of automation, the devaluation of data work is constructed through the logic of a new division of labour between model and data work, of pushing data work to its periphery and extracting more economic value through its devaluation and obfuscation. In a similar analysis of Amazon® Mechanical Turks (AMTs) under the concept of 'heteromation' (Ekbia and Nardi 2014, 2017), show how the well-defined computational tasks performed by human microworkers rendered their labour into 'bits of algorithmic functionaries' and helped in obscuring the very human labour that was being offered as a service on the platform. This novel way of dividing human-machine labour that ultimately hides instead of replacing the human and the value extraction from this very arrangement qualifies AMT as a 'heteromated' system, in their opinion (Ekbia and Nardi 2014).

The case of spatial organisation of data workers in *PriceWise* demonstrates a similar pattern of heteromation where data work that sustains the AI products and its value extraction in the company remains peripheral to its organisational hierarchy and becomes increasingly invisible through push for outsourcing to tier-2 cities. Second, even though analysts' jobs were neither precarious nor under direct threat of becoming redundant, the data workers did not envision these roles as desirable to their long-term career goals. Where the manager of the QA team at *PriceWise*

expected that the demand for their team's work would only grow, as the company ventured into newer domains and expanded its business, the young graduates working in the QA team understand the paradoxical position of their jobs within the organisation, particularly in the long term. They fear finding themselves in the sidelines of the AI economy with little prospect of upward mobility, rather than the technical role in the AI economy that they hoped for. Yet, they believe that working in a formal organisation that builds core AI products provides an opportunity for 'upskilling' and would increase their bargaining power in future job-searches, something gig work of a similar nature would not have provided.

Conclusion

Studies on data work that examine the role of the human worker and their work conditions emphasise the precariousness of these jobs, the instabilities of employment (Gray and Suri 2019; Posada 2022), invisibilisation of human workers through outsourcing or platformisation (Poster 2016; Gray and Suri 2019; Crain, Poster and Cherry 2016), and the harmful impact of the work itself (Roberts 2016). The case of *PriceWise* presented in this paper does not fit into these typical characterisations of data work within the AI ecosystem. Here, human data workers are part of an in-house data team with a fixed salary and their work does not have the kind of harmful impact that is, for instance, reported in content moderation studies. Moreover, these jobs are often advertised as new age service jobs created by the AI industry, sometimes even positioned as creating opportunities for marginalised communities such as the urban poor, rural youth, educated women or refugees (Janah 2017). Hence, the case of *PriceWise* allows us to examine the 'not so precarious' job roles of data work in the Global South and analyse how these different structural conditions nonetheless, are implicated in the displacement, devaluation and deskilling in data work. We argue that the problems of precarity and invisibility of data work are not inherent to AI or its specific technological design. Rather, they are produced through the specific embedding of AI production within the political economy of startup capitalism (Gupta 2019). Hence, mere formalisation of data work (as opposed to gigs) or optical visibility (Raval 2021) through regularisation of precarious work, does not address the structural challenges confronting the future of data workers. In our study, we find two simultaneous and intersecting processes of spatial and social displacement of data workers even within the formal sector employment, achieved through organisational arrangements of job roles associated with data work and codification of the work practices performed by these job roles. By drawing on earlier patterns of social and spatial displacement from the existing literature, our analysis shows the specific mechanisms of this displacement among formal sector data workers in India. By situating data work within a formal organisational set up, we underline how organisational policy of balancing investor versus client preferences shape work arrangements and spatial division of the data workers, how the relations between data and 'core' technical teams helps to invisibilise the human labour in the production of AI and how increasing codification of data work, leads to devaluation of data work as 'menial work' within the organisation and deskilling of young data workers at large, making them vulnerable in choosing a meaningful career path of their choice.

The issue of this large-scale deskilling in formal employment needs closer scholarly attention as it pertains to worrying trends of skill and employment trends in India and the Global South, with implications beyond AI and work. While skill development of the youth has received much push from the Indian state in recent times, general trends of employment seems to be focussed upon providing the industry 'with a steady labour supply than enhancing quality of employment, working standards or wage levels' (Carswell and De Neve 2024, 2). Amidst growing unemployment in India (Dugal 2023), a pragmatic approach of catering to the immediate demands of the global AI industry with cheap and flexible labour force takes precedence over nurturing specific skills development (Carswell and De Neve 2024). As AI is deployed in ever newer domains, data work is only expected to increase, rendering it a major source of employment within the AI industry. Moreover, the prevalence of outsourcing within the industry will make these jobs more desirable to the young workforce in the Global South. Against this backdrop, problematizing this new form of work primarily through the analytical frames of invisibility and precarity, while undoubtedly important, may not sufficiently grapple with the long-term impact on skill and career trajectories of the young and trained workforce in Global South, engendering massive deskilling of human labour through AI. Hence, through this study, we seek to reorient the discourse of AI and the future-of-work, to pay close attention to the current conditions of labour in the production of AI and the future of the workers engaged in it. We do so on three accounts: first, we challenge the tech-deterministic view of automation where AI assumes subject position, making it appear more as an inevitable moment in history of automation (Shestakofsky 2017, 2020); second, we resist the projection of AI as disruptive technology, as it drifts our attention from the specific trajectories of automation and labour process rooted in a time and space that ultimately anchors the production of AI; and lastly, we question the narrow view of human's position vis-a-vis the AI, that builds on inevitability, and through that, seek to recenter the conversations on AI and future-of-work around the specific labour processes of AI and its implications on human workers.

Acknowledgements

Author Biographies

Bidisha Chaudhuri is an Assistant Professor of Government, Information Cultures and Digital Citizenship at the University of Amsterdam and is interested in digital governance and

infrastructures, politics of data and algorithms, political economy of digital technologies and work practices and AI ethics.

Srravya Chandhiramowuli is a PhD candidate at the University of Edinburgh's Institute for Design Informatics, examining data annotation labours within the global supply chains of AI production.

References

Akst, Daniel. 2013. "Automation Anxiety." *The Wilson Quarterly* 37(3). https://www.jstor.org/stable/wilsonq.37.3.06.

Aneesh, Aneesh. 2001. "Skill Saturation: Rationalization and Post-Industrial Work." *Theory and Society* 30(3): 363–96. https://www.jstor.org/stable/657966.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118(4): 1279–1333. https://doi.org/10.1162/003355303322552801.

Baum, Tom. 2008. "The Social Construction of Skills: A Hospitality Sector Perspective." *European Journal of Vocational Training* 44(2): 74–88. Accessed June 13, 2024. https://www.cedefop.europa.eu/files/etv/Upload/Information_resources/Bookshop/503/44_en_Baum.pdf.

Bilić, Paško. 2016. "Search Algorithms, Hidden Labour and Information Control." *Big Data & Society* 3(1): 1–9. https://doi.org/10.1177/2053951716652159.

Braverman, Harry. 1974. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century.* New York: NYU Press. https://www.jstor.org/stable/j.ctt9qfrkf.

Carswell, Grace, and Geert De Neve. 2024. "Training *for* Employment or Skilling up *from* Employment? Jobs and Skills Acquisition in the Tiruppur Textile Region, India." *Third World Quarterly* 45(4): 715-733. https://doi.org/10.1080/01436597.2022.2156855.

Chandhiramowuli, Srravya, and Bidisha Chaudhuri. 2023. "Match Made by Humans: A Critical Enquiry into Human-Machine Configurations in Data Labelling." In *Proceedings of the 56th Hawaii International Conference on System Sciences*, 2007–2016. https://hdl.handle.net/10125/102882.

Charmaz, Kathy. 2017. "The Power of Constructivist Grounded Theory for Critical Inquiry." *Qualitative Inquiry* 23(1): 34–45. https://doi.org/10.1177/1077800416657105.

Chernoff, Michael. [2010] 2013. "Social Displacement in a Renovating Neighborhood's Commercial District: Atlanta." In *The Gentrification Debates: A Reader*, edited by Japonica Brown-Saracino 295. New York: Routledge.

Cockburn, Cynthia. [1985] 1999. "'Caught in the Wheels: The High Cost of Being a Female Cog in the Male Machinery of Technology' and 'The Material of Male Power.'" In *The Social Shaping of Technology*, edited by Donald Mackenzie and Judy Wajcman. Open University Press.

Cole, Matthew, Hugo Radice, and Charles Umney. 2021. "The Political Economy of Datafication and Work: A New Digital Taylorism?" *Socialist Register* 57: 78-99.

Crain, Marion, Winifred Poster, and Miriam Cherry, eds. 2016. *Invisible Labor: Hidden Work in the Contemporary World*. Berkeley: University of California Press. https://socialistregister.com/index.php/srv/article/view/34948/26740.

Dugal, Ira. 2023. "Where Are the Jobs? India's World-beating Growth Falls Short." *Reuters*, May 31, 2023. Accessed June 20, 2024. https://www.reuters.com/world/india/despite-world-beating-growth-indias-lack-jobs-threatens-its-young-2023-05-30/.

Silkin, Lewis. 2021. "Deskilling: What Are the Historical, Societal and Legal Implications? — Future of Work Hub." *Futureofworkhub* blog. April 28, 2021. Accessed June 15, 2024. https://www.futureofworkhub.info/explainers/2021/4/28/deskilling-what-are-the-historical-societal-and-legal-implications.

Ekbia, Hamid, and Bonnie Nardi. 2014. "Heteromation and Its (Dis)Contents: The Invisible Division of Labor between Humans and Machines." *First Monday* 19(6). https://doi.org/10.5210/fm.v19i6.5331.

———. 2017. *Heteromation, and Other Stories of Computing and Capitalism*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/10767.001.0001.

Graham, Mark, Isis Hjorth, and Vili Lehdonvirta. 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): 135–62. https://doi.org/10.1177/1024258916687250.

Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Eamon Dolan Books.

Gupta, Hemangini. 2019. "Testing the Future: Gender and Technocapitalism in Start-up India." *Feminist Review* 123(1): 74–88. https://doi.org/10.1177/0141778919879740.

Hicks, Mar. 2017. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: MIT Press.

Hube, Christoph, Besnik Fetahu, and Ujwal Gadiraju. 2019. "Understanding and Mitigating Worker Biases in the Crowdsourced Collection of Subjective Judgments." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 407, 1–12. https://doi.org/10.1145/3290605.3300637.

Ilavarasan, P. Vigneswara. 2008. "Software Work in India: A Labour Process View." In *In an Outpost of the Global Economy: Work and Workers in India's Information Technology Industry*, edited by Carol Upadhya and A. R. Vasavi, 162–89. New Delhi: Routledge.

Irani, Lilly C. 2015a. "The Cultural Work of Microwork." *New Media & Society* 17(5): 720–39. https://doi.org/10.1177/1461444483511926.

———. 2015b. "Justice for 'Data Janitors.'" *Public Books Blog.* January 15, 2015. Accessed June 15, 2024. https://www.publicbooks.org/justice-for-data-janitors/.

Irani, Lilly C., 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*, 611–20. Paris, France: ACM. https://doi.org/10.1145/2470654.2470742.

Janah, Leila. 2017. *Give Work: Reversing Poverty One Job at a Time.* New York, Penguin Random House Publishing.

Joshi, Sonam. 2019. "How Artificial Intelligence Is Creating Jobs in India, Not Just Stealing Them." *The Times of India*, September 9, 2019. Accessed June 15, 2024. https://timesofindia.indiatimes.com/india/how-artificial-intelligence-is-creating-jobs-in-india-not-just-stealing-them/articleshow/71030863.cms.

Mehrotra, Karishma. 2022. "Human Touch." *Fifty-Two* blog. July 23, 2022. https://fiftytwo.in/story/human-touch/.

Miceli, Milagros, 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*, CSCW 4(2): 115, 1-25. https://doi.org/10.1145/3415186.

Miceli, Milagros, and Julian Posada. 2022. "The Data-Production Dispositif." *Proceedings of the ACM on Human-Computer Interaction* 6(2): 460, 1-37. https://doi.org/10.1145/3555561.

Munn, Luke. 2022. *Automation Is a Myth*. Redwood City, CA: Stanford University Press.

Murali, Anand. 2019. "How India's Data Labellers Are Powering the Global AI Race." *FactorDaily*, March 21, 2019. Accessed June 14, 2024. https://archive.factordaily.com/indian-data-labellers-powering-the-global-ai-race/.

Murgia, Madhumita. 2019. "AI's New Workforce: The Data-Labelling Industry Spreads Globally." *Financial Times Blog.* August 8, 2019. Accessed June 14, 2024. https://medium.com/financial-times/ais-new-workforce-the-data-labelling-industry-spreads-globally-f472cb1bac09.

Murphy, Michelle. 2017. *The Economization of Life*. Durham, NC: Duke University Press.

Natarajan, Sarayu, Suha Mohamed, Kushang Mishra, and Alex Taylor. 2021. "Just and Equitable Data Labelling: Towards a Responsible AI Supply Chain." Report. *Aapti Institute.* February 25, 2021. Accessed June 14, 2024. https://aapti.in/blog/just-and-equitable-data-labelling/.

Newlands, Gemma. 2021. "Lifting the Curtain: Strategic Visibility of Human Labour in AI-as-a-Service." *Big Data & Society* 8(1):1–14. https://doi.org/10.1177/20539517211016026.

Pietrobelli, Carlo, and Roberta Rabellotti. 2011. "Global Value Chains Meet Innovation Systems: Are there Learning Opportunities for Developing Countries?" *World Development* 39(7): 1261–69. https://doi.org/10.1016/j.worlddev.2010.05.013.

Posada, Julian. 2021. "Why AI Needs Ethics from Below." *AI Now Institute Website*, Guest Post in the entry "Labor" of the "New AI Lexicon." September 23, 2021. Accessed June 14, 2024. https://ainowinstitute.org/publication/a-new-ai-lexicon-labor.

———. 2022. "Embedded Reproduction in Platform Data Work." *Information, Communication & Society* 25(6): 816–834. https://doi.org/10.1080/1369118X.2022.2049849.

Poster, Winifred R. 2016. "The Virtual Receptionist with a Human Touch: Opposing Pressures of Digital Automation and Outsourcing in Interactive Services." In *Invisible Labor: Hidden Work in the Contemporary World*, edited by Marion Crain, Winifred Poster, and Miriam Cherry, 87–112. Berkeley: University of California Press. https://doi.org/10.1525/9780520961630-007.

Raval, Noopur. 2021. "Interrupting Invisibility in a Global World." *ACM Interactions* 28(4): 27–31. https://doi.org/10.1145/3469257.

Rigby, Mike, and Enric Sanchis. 2006. "The Concept of Skill and Its Social Construction." *European Journal of Vocational Training* 37: 22–33. Accessed June 15, 2024. https://www.cedefop.europa.eu/files/etv/Upload/Information_resources/Bookshop/430/37_en_rigby.pdf.

Roberts, Sarah T. 2016. "Commercial Content Moderation: Digital Laborers' Dirty Work." In The Intersectional Internet: Race, Sex, Class, and Culture Online, edited by Safyia Umoja Noble and Brendesha M. Tynes, 147–159.

Sambasivan, Nithya, Shivani Kapania, Hannah Highfill, Diana Akrong, et al. 2021. "'Everyone Wants to Do the Model Work, Not the Data Work': Data Cascades in High-Stakes AI." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 39: 1–15. https://doi.org/10.1145/3411764.3445518.

Sanyal, Kalyan, and Rajesh Bhattacharyya. 2009. "Beyond the Factory: Globalisation, Informalisation of Production and the New Locations of Labour." *Economic and Political Weekly* 44(22): 35–44. https://www.jstor.org/stable/40279056.

Shestakofsky, Benjamin. 2017. "Working Algorithms: Software Automation and the Future of Work." *Work and Occupations* 44(4): 376-423. https://doi.org/10.1177/0730888417726119.

———. 2020. "Stepping Back to Move Forward: Centering Capital in Discussions of Technology and the Future of Work." *Communication and the Public* 5(3–4): 129–33. https://doi.org/10.1177/2057047320959854.

Tubaro, Paola, 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).
https://doi.org/10.1177/2053951720919776.

Upadhya, Carol. [2011] 2020. "Software and the 'New' Middle Class in the 'New India.'" In *Elite and Everyman: The Cultural Politics of the Indian Middle Classes*, edited by Amita Baviskar and Raka Ray, 167–92. New Delhi: Routledge.

Upadhya, Carol, and Aninhalli Rame Vasavi, eds. 2008. *In an Outpost of the Global Economy: Work and Workers in India's Information Technology Industry*. New Delhi: Routledge India.

Wang, Ding, Shantanu Prabhat, and Nithya Sambasivan. 2022. "Whose AI Dream? In Search of the Aspiration in Data Annotation." In Proceedings of the 2022 *CHI Conference on Human Factors in Computing Systems* 582: 1–16. https://doi.org/10.1145/3491102.3502121.

Warhurst, Chris, Chris Tilly, and Mary Gatta. 2017. "A New Social Construction of Skill." In *The Oxford Handbook of Skills and Training*, edited by John Buchanan, David Finegold, Ken Mayhew, and Chris Warhurst, 72–91. https://doi.org/10.1093/ofordhb/9780199655366.013.4.