The Deskilling of Domain Expertise in AI Development

Nithya Sambasivan
nithyas@gmail.com
Unaffiliated
San Francisco, CA, USA

Rajesh Veeraraghavan
rajesh.veera@georgetown.edu
Georgetown University
D.C., USA

ABSTRACT

Field workers, like farmers and radiologists, play a crucial role in dataset collection for AI models in low-resource settings. However, we know little about how field workers' *expertise* is leveraged in dataset and model development. Based on 68 interviews with AI developers building for low-resource contexts, we find that developers reduced field workers to data collectors. Attributing poor data quality to worker practices, developers conceived of workers as corrupt, lazy, non-compliant, and as datasets themselves, pursuing surveillance and gamification to discipline workers to collect better quality data. Even though models sought to emulate the expertise of field workers, AI developers treated workers as non-essential and *deskilled* their expertise in service of building machine intelligence. We make the case for why field workers should be recognised as domain experts and re-imagine domain expertise as an essential partnership for AI development.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

ML, labour studies, deskilling, Taylorism, field workers, data quality, data collection, domain expertise

ACM Reference Format:
Nithya Sambasivan and Rajesh Veeraraghavan. 2022. The Deskilling of Domain Expertise in AI Development. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3491102.3517578

1 INTRODUCTION

Breakthroughs in AI model performance and capabilities have led to a palpable interest in deploying machine learning in low-resource contexts, where they are seen as potential opportunities to address socio-economic challenges. AI models often scale up the expertise of domain experts that may be unevenly distributed due to infrastructural and economic challenges, *e.g.,* scaling instant Tuberculosis reads in regions where there is only one doctor per 30,000 people.

Noteworthy as their objectives and efforts are; in reality, several AI projects in low-resource areas stall as research experiments, rather than as sustained real-world deployments. Poor quality data that is unusable, incomplete, or inaccurate for training purposes is one of the fundamental hurdles to building and deploying models [44, 76, 85]. Readily available datasets are often lacking in low-resource areas, in contrast to academic or enterprise settings [44]. To address this lack of data problem, developers rely on the labour of field workers, such as agricultural extension workers and community health workers, to create AI/ML datasets to train models [76]. AI developers in low-resource contexts often take on bootstrapping and maintenance labour that is under-appreciated in conventional AI research, such as partnering with institutions and working with infrastructural constraints [76]. While AI developers rely on several human intermediaries to build the datasets, we do not yet understand how they contend with these human intermediaries.

The disregard for domain expertise in model building is well-known in AI, *e.g.,* [17]. Statements like "*every time I fire a linguist, the performance of the speech recognizer goes up*" (attributed to NLP researcher Frederick Jelenik) [35] reflect how human experts are not fully integrated and recognised as a part of the technical AI pipeline. More generally, it is indeed a well-recognised problem that computer scientists and engineers often do not recognise the expertise of workers, despite their crucial contributions [34, 79]. AI models in low-resource areas are highly dependant upon the critical labour of field workers for dataset creation and deployments. Despite their skills and domain expertise that takes years to build, field workers have been reported to perform low-waged, arduous, and burdensome work for AI data collection [41, 56]. A wider body of work in Crowdwork and Human Computation has focused on centering the invisible AI/ML labourer, such as annotators and content moderators [15, 21, 30, 31, 39, 52]. In contrast to these previous studies, where the focus is on visibilizing the labour of computational workers who are directly recruited and compensated for their work (albeit often inadequately), our paper focuses on field workers who are recruited indirectly and invisibly through partner organizations for AI dataset collection, often for no additional compensation for dataset collection. The field workers typically take decades to build expertise in their respective socio-economic areas, such as rearing crops, delivering health care, protecting endangered species, and so on, with deeply embedded relations in their communities and phenomena. Importantly, we empirically know little about how this hard-earned *expertise* of field workers is seen by the AI practitioners who build and deploy these systems. Given that models seek to emulate and aggregate local expertise, to what extent is the worker's expertise acknowledged, engaged with, or credited by the AI developer? Such an endeavour goes beyond mere inclusion of labourers, and into examining the particular ways in which labourers fit into the mental models of practitioners, who exert power and make consequential decisions while building AI models.

In this paper, we fill this crucial gap by 'studying up' how AI developers conceive of and manage the expertise of field workers

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517578

in dataset and model building. We draw from 68 interviews with AI practitioners building models for low-resource communities in India, Sub-Saharan Africa, and the United States. We find that developers viewed field workers only as data collectors, and disregarded their field expertise. Importantly, we find that AI developers attributed poor data quality to poor work practices of field workers. We present four conceptions of field workers as held by developers: workers as corrupt, lazy, non-compliant, and as the dataset itself. The vast majority of developers did not appear to compensate field workers for surplus data labour, provide disclosure or training on models or downstream use cases that workers were contributing towards, and meaningfully attribute credit to field workers for their contributions to model outcomes. Despite the limited engagement and understanding of field workers, developers reported creating several disciplinary interventions to influence field workers to collect better quality data, in the form of surveillance, gamification, cross-verification, and pre-processing fixes. Our focus is not on verifying the validity of developer views of field workers, but on reporting the existence and prevalence of these views, which are sometimes actioned upon as punitive interventions that can be detrimental to the well-being of field workers. We also found that even though a few AI developers were reflexive, they felt constrained and attested to their own lack of agency in addressing the limits of the AI pipeline.

We argue that field workers should be accurately viewed as domain experts with special knowledge and mastery in their local contexts. The AI development apparatus in low-resource areas deskills[1] and invisibilises the domain expertise of field workers. Even though the models in our study sought to emulate and improve over the expertise of field workers, their expertise was treated as non-essential knowledge by AI developers. The worker's knowledge, networks, and capabilities were viewed as assets for dataset collection, rather than necessary expertise for system building. Developers held reductionist assumptions about the expertise of field workers, enrolling them as invisible data collectors that most did not have any direct contact with. AI developers, who were experts in their technical fields but not in the application-domains, *e.g.,* a developer building a cancer prediction model, will inevitably leave gaps when domain experts are excluded from the model that seeks to learn their knowledge (through datasets). But if we consider the domain expertise of field workers as an essential partnership throughout the AI pipeline, we can see new possibilities for collecting, modelling, and scaling knowledge. Domain experts can contribute to critical questions that can impact model behaviours: What exactly are we modelling? What assumptions are appropriate? What features should be included in the model? What are we trying to predict? How will we know? While machine intelligence could be more effective, accurate, or faster at the work, the domain experts are not always consulted enough to be able to evaluate such claims. When we recognise field workers as domain experts in the first place, the problem of poor quality data can then be read differently as contingent and practical challenges in workflows of data collection, and the design solutions might be oriented towards solving these practical problems, rather than manifesting as extractive and punitive solutions.

Our discussion re-imagines how we can shift to viewing field workers as full-range domain experts in their communities and contexts. We identify pragmatic ways to co-create datasets, algorithms, and evaluation metrics with domain experts. We call for reflection on normative assumptions in AI, such as how the AI/ML pipeline is imagined, and whose expertise counts. We press upon AI developers the need to move away from control and influence of field workers to collect better or more data, to embrace more participatory stances. We present opportunities to recognise the untapped and hidden labour of field workers in AI development.

Our paper makes two main contributions. First, we present empirical evidence for how AI developers building models for low-resource areas do not see the domain expertise of field workers that power their datasets, based on their limited conceptions of workers and the disciplinary interventions they create to manage data collection practices. We make the case that AI development in low-resource areas deskills the expertise of field workers involved. Second, we present various implications for recognising, respecting, and rewarding domain expertise in the development of machine intelligence in low-resource areas.

2 RELATED WORK

2.1 Deskilling and expertise in automation

The workers of ML and AI are divided between the "coding elite" [10] who do the innovative work and the "cybertariat" [36], who does the "menial work" of building data that feed or train the algorithms [38]. This classificatory struggle of what constitutes expertise has social origins with material consequences. Labour theorists have blamed the rise of capitalism with the associated phenomenon of "Taylorism" and the use of novel technologies of production in *deskilling labour.* ultimately resulting in shifting power to capital over labour [7]. Deskilling is a reduction in the skills and knowledge needed to do a job when automation degrades human work. Others have argued taking a long term view beginning with preindustrialism and ending with AI, technology has played a positive as well as a negative role [25]. The extent of deskilling's scope in explaining capitalist development has been criticized extensively [4].

However, to the extent automation in work processes results in integrating knowledge into labour displacing machines, a deskilling occurs with workers being further relegated to menial tasks[96]. As a result, to make workers accountable, strategies to build better pricing mechanisms [77], and increasingly surveillance and control of workers is often seen as necessary by management[96] [47] [86] which have historically met with conflict and worker resistance [9]. Information and data-based work processes that include work on AI, makes work visible to employers that Zuboff calls "informating" [96]. Automating these process have also allowed for tools to be built to empower workers [40] [70]. Another strand of research focuses on forecasting of job losses from computerised automation [33, 48, 65]. Frey and Osborne predicted professions like lawyers,

[1]Deskilling, first coined by sociologist Harry Braverman in 1974, occurs when sophisticated technologies are introduced that no longer require workers to have formerly necessary skills [7]. Deskilling of work has been observed throughout the history of technology and industry, notably among professions like skilled craftsmen [88], nurses [68] and farmers [24]; deskilling has also been discussed in the context of job losses from automation [26]

doctors and accountants were being broken down into their component parts via the increasing use of algorithms allowed apps to create online self-help services and provide a diagnosis [26].

James C. Scott used the Greek phrase mẽtis for local expert knowledge, to represent "a wide array of practical skills and acquired intelligence responding to a constantly changing natural and human environment."[74] Lave and Wenger located expertise not in local knowledge but in *communities of practice* through "legitimate peripheral participation" [45]. Local experts develop expertise through a "shared repertoire of resources: experiences, stories, tools, ways of addressing recurring problems—this takes time and sustained interaction"[90]. While Scott was focused on how "high modernist" actions fail because of the impossibility of tapping into mẽtis, Denton was optimistic about the possibility of algorithms tapping into local expertise, he writes, "a trained neural network is still essentially a mẽtis solution, turning a body of experience (training data) into a specialized knowledge of a problem domain[2]. But the local knowledge of image classification has been offloaded from the experts to the machines. This is a real shift"[16]. But this transfer of expertise from humans to machines is not straightforward. Ackerman et al. review the CSCW literature on organizational knowledge sharing and explore the shift from the repository model of knowledge sharing to tacit knowledge sharing[1]. Studies that have examined knowledge sharing in 'expert systems' has recognised the importance of socio-political considerations [66] [20] as well as the limits of knowledge sharing between communities of practice [19].

We take a step back and critically examine the labour process of building expertise in these algorithmic systems. In particular, we focus on the politics of recognition of expertise and its consequences. Our paper makes the case for how and why deskilling occurs in AI development, and how the expertise of field workers is reduced by typecasting them as data collectors in service of building uber-expert models. Our work adds to the focus on AI labour practice by turning the gaze on the "coding elite" and examining their perspectives.

2.2 Invisible computational labour

Under-recognized and invisible labour is a persistent challenge in computing, widely studied in the context of technology-mediated work situations [54, 78, 79]. The first computers were human beings, often women, who were often rendered invisible [32, 34]. Recent studies have pointed to the invisible agency of humans or "ghost work" [31] in algorithmic automation [61]. Critical data scholars have shown that data is never "raw" [29] and have shown how undervalued human labour, powers these hyper visible AI models with a focus skewed towards algorithmic work (*e.g.,* heteromation ([21]; fauxtomation ([82]). Recent scholarship has examined how frontline health workers in India navigate data collection amidst the multiple demands placed on them [41, 58, 83] and the tensions that arise across workers and developers, calling for greater attention to worker agency, comfort levels, and training [58], transparency and accountability [83], and pointing to how solidarity is practiced among the workers themselves in advocating for marginalised groups [41]. Prior work has also examined how mental models of AI intermediaries who work with marginalized citizens critically impact outcomes [56].

Researchers point to the need to bring the labourers back in and to critically examine the labour process in producing AI models [15, 21, 31, 39, 52, 89]. Examining data work is critical not just for its intrinsic value, but also for downstream "data cascades" as introduced by Sambasivan *et al.* [72] which negatively impact model outcomes. Data work is also important for injecting flexibility in "street-level algorithms" to overcome the limits of algorithmic expertise [3] as well as to more deeply integrate into prioritizing local ethics as opposed to relying on statistical rationality [81] .

Studies have sought to recognize the complexity and intricacies involved in building datasets. For example, Møller *et al.* [52] discuss how the data work of clerical hospital workers is complex and requires a significant amount of domain expertise in navigating ethical quandries [52]. Studies have shown how this data labour is often not recognized (the work on Mechanical Turkers by Martin *et al.* [49], data annotators by Wang *et al.* [89], and frontline health workers by Ismail and Kumar [41]) and Thakkar *et al.* [83]).

It is important to understand why domain expertise is not recognized by AI developers. Suchman showed how engineers underestimated the difficulties of what the secretaries did [80]. Our research fills a crucial gap of detailing the perspectives of the developers of AI models, and how they imagine and characterise expertise, recruit labourers, and manage their labour to build models.

2.3 Poor quality data in low-resource areas

Data access and quality is a significant challenge in low-resource contexts, due to the lack of critical human and institutional capacity for creating modern data pipelines [60] [14, 63][8]. An active area of research is focused on coping with limited datasets through technical solutions, including dimensionality reduction [95] and data augmentation [22].

It is standard practice for AI developers to work with local field partners to create datasets [44, 76, 85] building on the past two decades of ICTD, GIS, and other data-driven systems building in low-resource areas (*e.g.,* see [12, 18, 27, 60, 63]. The onus of data collection thus falls on local experts, such as community health workers and nurses, whose primary jobs, such as providing patient care, are often orthogonal to dataset collection. Even so, such local experts are required to perform these tasks anyway [72]. Prior research has documented how dataset collection exacerbates the lack of agency of field workers [41], done with inadequate training [12, 55, 58, 63], or mismatched incentives [12, 72, 87].

Dealing with limited datasets in low-resource areas is challenging, as training datasets can be small, biased, and non-generalisable [76], with practitioners in these settings often running into inconsistent, poorly recorded, incorrectly formatted, missing, and fabricated data [62, 63, 76]. Additionally, factors like competing organisational agendas, misaligned dataset needs for program management vs. model building, changing requirements, lack of structured forms, and the heavy reliance on paper in low-resource areas can negatively impact the quality of datasets [14, 28, 44, 63]. Our paper establishes a direct, causal, and active link held by developers that workers are to be blamed for poor data quality, through their conceptions of labourers and interventions they create to control the

[2]Thanks to Meghan Mandi for pointing this connection out

work of workers. Our research provides evidence for the widespread nature of these beliefs in our globally-distributed sample of AI developers.

3 METHOD

Our research methods were chosen to provide a detailed understanding of the goals, labour enrolment, and labour processes of AI developers building for various low-resource contexts around the world. We "studied up" [53] AI developers and researchers to gain a nuanced picture of their animating visions and the ground realities of model development, from their vantage points as creators and builders of the systems. Consequentially, the scope of our study is limited to accounts from system builders, not field workers (see Limitations). AI developers provided us with reflexive accounts of their own roles, choice of algorithms, and labour arrangements. We empirically report on the viewpoints and practices of developers in working on AI projects that enrolled field workers. Our research used a qualitative approach of semi-structured interviews and an inductive analysis. Our interview data collection was done in two phases; the first phase between May and July 2020, and the second phase between April and August 2021, with analysis and synthesis done in between and after. We conducted interviews with 68 AI developers working in low-resource areas.

As a disclosure of postionality, we do not seek to endorse the claims and views of these AI developers, but are concerned with understanding their perceptions and strategies with field workers, as builders of these AI systems. As we make explicit in our Introduction and Discussion, the authors' perspectives are critical of views and practices that largely dehumanise field workers and reduce them to invisible data labourers. We go on to provide practical, structural and intellectual suggestions for recognizing the domain experts of field workers throughout the AI pipeline.

Recruitment and moderation. We recruited participants through a combination of developer communities, distribution lists, professional networks, and personal contacts, using snowball and purposive sampling [59] that was iterative until saturation. All participants were directly involved in building models for low-resource contexts. We limited our sample to AI applications in high stakes applications, a common pairing in low-resource settings. Participants played technical roles, as engineering leads, founders, developers, and program managers, and were directly involved in developing systems. Interviews were conducted using video conferencing, due to COVID-19 travel limitations. These interviews were focused on identifying project visions, understanding the data pipelines, labour enrolment and processes, data quality challenges, challenges and approaches in low-resource development, developer conceptions of labour, value conflicts, and interventions. We conducted all interviews in English (preferred language of participants). Each participant received a thank you gift in the form of a gift card. Amounts were localised in consultation with regional experts, based on purchasing power parity and non-coercion (100 USD for the US, 27 USD for India, 35 USD for East and West African countries). Due to workplace restrictions, we were not able to compensate government employees. Each session focused on the participant's experiences, practices, and challenges in AI development and lasted 60 to 75 minutes each. Interview notes were recorded in the form of field notes or video recordings, transcribed within 24 hours of each interview.

We intentionally sampled AI developers building for low-resource communities in various countries, in order to gain diversity of experiences and contextual factors that affect data quality. AI developers were located in, or worked primarily on projects based in, India (32), the US (22), Nigeria (10), Kenya (2), Uganda (1), and Ghana (1). We interviewed 55 male and 13 female AI developers. Participants were employed in startups (28), large companies (24), academia (12), and non-profits (4). All projects depended on field workers enrolment for dataset collection and labelling; 40 projects depended priarily on data collected by field workers solely, where 30 projects also utilised sensor data, end-user data, open datasets, crowd-sourced and client data. Examples of field workers include community health workers, agricultural extension workers, radiologists, and truck drivers. Refer to Table 1 for details on participant demographics.

Co-authors of this paper have partnered with marginalised communities in technology design for nearly thirty years. Our positionalities are shaped by our disciplinary commitments to HCI, ICTD, and Technology Policy.

Analysis. Using inductive analysis [84], the two co-authors independently read all units multiple times, and categories (unit of analysis) were initially identified by each researcher, together with a description and examples of each category, until a saturation point was reached. Our upper level categories were guided by the evaluation aims, comprising (1) project vision and problem selection; (2) trade-offs with using AI; (3) defining the right dataset; (4) human factors in dataset collection; (5) labour quantity and quality enrolled; (6) developer involvement of experts; (7) views on field workers; (8) strategies and interventions to manage data quality; (9) challenges in low-resource AI development; and (10) reflexive accounts on blindspots and retrospectives. The categories were iteratively refined through group discussions with meeting, diverging, and synthesizing during the analysis phase. Further iterations resulted in the formation of lower-level categories such as "strategies: monitoring". These categories were consolidated into two top-level categories of developer conceptions of field workers and interventions created, as well as 14 nested categories, which included skill, partnership, and incentives.

Research ethics. We took intentional care to create a research ethics protocol to protect respondent privacy and safety, especially due to the sensitive nature of our inquiry. During recruitment, participants were informed of the purpose of the study, the question categories, and researcher affiliations. Participants signed informed consent acknowledging their awareness of the study purpose and researcher affiliation prior to the interview. At the beginning of each interview, the moderator additionally obtained verbal consent. We stored all data in a private Google Drive folder, with access limited to our team. To protect participant identity, we deleted all personally identifiable information in research files. We redacted identifiable details when quoting participants. We acknowledge that the above is our interpretation of research ethics, which may not be universal.

Limitations and future work. Our work is a snapshot of AI designers and their perspectives through semi-structured interviews. To fully understand the extent of deskilling, we would need

Type	Count
Roles	AI Engineer **(19)**, Startup Founder **(18)**, Professor **(12)**, Data Scientist **(8)**, Research Scientist **(7)**, Program Manager **(5)**
Location	India **(32)**, US **(22)**, Nigeria **(10)**, Kenya **(2)**, Ghana **(1)**, Uganda **(1)**
Gender	Male **(55)**, Female **(13)**
Organisation	Startup **(28)**, Large company **(24)**, Academia **(12)**, Non-profits **(4)**
AI application	Health **(27)** , Agriculture **(11)**, Climate and environment **(10)**, Finance **(7)**, Public safety **(4)**, Wildlife conservation **(2)**, Aquaculture **(2)**, Education **(2)**, Employment **(1)**, Robotics **(1)**, Fairness in ML **(1)**
Field workers	Health workers (nurses, community health workers, radiologists, oncologists, gynaecologists, general practitioners, machine operators): **(23)**, Agricultural workers (agricultural extension workers, farmers, district-level scientists): **(5)**, Ecological and environmental workers (wildlife patrollers, oceanologists): **(7)**, Transportation workers (truck drivers, boda-boda drivers): **(2)**, Education workers (teachers, teaching assistants): **(1)**, Data from additional sources (sensor data, open data, user data, pre-existing data, crowd-sourcing): **(30)**

Table 1: Summary of participant demographics

to examine work practices over a period time and by direct observation. Our work can also be extended by interviewing field workers and intermediaries to understand their perspectives on the challenges of data and model building. Our paper's focus was to understand the perceptions of how data workers were perceived by developers. It should be clear that our work does not concern itself with verifying the veracity of the judgements of the AI developers; rather, our focus is on reporting the very existence of these views and biases, and subsequent enactment of disciplinary interventions on field workers by developers, despite the seminal contributions of workers to AI systems (often for free or with limited consent and transparency). Our concern is to understand the mental models of how developers view these essential field workers in their algorithm development. Future work can address this research limitation by triangulating and understanding the experiences of field workers. However, we believe that solely understanding the mental models of the developers is a standalone contribution. Future work can include systematically measuring the various effects of including domain expertise on model development and deployment. Another area for future research is to study the processes of inclusion of human intermediaries for AI projects in well-resourced areas.

All interviews and analysis were conducted over video and phone, due to the COVID-19 pandemic. As a result of travel restrictions, we were unable to conduct any observational research of field workers interactions that would have otherwise been possible. However, we feel that the self-reported conceptions and interventions have validity, and sufficient rigour and care was applied in covering the themes through multiple questions and solicitation of examples. Gender distribution in our study is reflective of the AI industry's disparities [91] and sampling limitations.

4 FINDINGS

All AI projects in our study aimed to use models to scale up deficits of resources, in various social, economic, and ecological domains. Developers typically described their visions of introducing algorithms with an efficiency function of improved time, speed, or accuracy, or reduced cost when compared to the status quo of humans. All project visions were motivated by how models could make services instantly and widely available, considering the specific deficits of low-resource contexts. For example, models were seen as impactful interventions that could recognise patterns which were previously limited to only some experts, identify critical conditions in a few seconds instead of the status-quo of weeks, and identify high-risk targets to intervene on which cognitively overburdened field workers were unable to identify. Domain expertise was viewed by developers as a scarce, limited resource which models could train on and scale up. As P23 said, "*you're trying to transition all the knowledge from a human brain [agricultural worker] to machine so that you can start to have a very good repeatability of how somebody does the procurement of the produce.*" In some cases, models were justified based on anecdotal or small data, despite the fact that models require moderate to high scale operations of labour and infrastructure, *e.g.,* P59 spoke of their crop disease model, "*we did a survey of 25-30 field workers and only 2 could identify pests.*" Once the projects were initiated, the field worker played an important role in creating the datasets for training models. Below, we classify how AI developers perceived and discussed field workers.

4.1 Reduction of field workers to data collectors

Field workers had existing responsibilities, and dataset collection was often orthogonal and surplus labour, except in five cases where workers were recruited independently and compensated. Developers entered into partnerships with field organisations, mostly governmental or non-profit institutions, to provide field and logistical support for model building and deployment. The field workers were typically employees or sub-contractors of the partner organisations in pre-existing roles, *e.g.,* school teachers, community health workers, and agricultural extension workers. field workers were auto-enrolled into data collection responsibilities, but these were double duties as primary objectives like rearing crops, delivering healthcare, or protecting endangered species still needed to be fulfilled. (The arrangement we report of is in contrast to crowd work and human computation studies, where computational workers are directly recruited and compensated, however inadequately, for AI/ML dataset tasks *e.g.,* [13].) Field workers in our study were health workers, nurses and doctors; farmers; school teachers; forest guards; diagnostic machine operators; taxi drivers; traffic police; oceanologists; and voluntary, pro-bono contributors. In nearly all cases (barring five), developers did not report compensating labourers; in all cases, the work of dataset collection was supplementary to the labourer's primary work.

Despite the direct impact of a field workers expertise and work practices on data quality, developers normatively did not engage with field workers in any capacity, including field visits during data collection or soliciting their inputs on systems. Most developers spoke of field workers using language of low-level computational tasks, such as 'counting', 'clicking', 'sorting', 'recording', or 'collecting'. Developers often spoke of the need for building good relationships with field workers, but pointed to time, budget, or organisational constraints as obstacles coming in the way of having direct interactions with field workers and recognising their expertise. As P54 explained, "*We end up talking to very senior people of the project who know a lot, but not the person creating the dataset. It would have been helpful to get from the health workers directly. That never happens.*" Only in a handful of cases (refer to Engagement under Strategies) did developers directly engage with field workers and solicit their knowledge and expertise, train them, or engage them in problem specification. Contrarily, a vast majority of developers interfaced and built relationships with the leaders and senior staff of the partner organisations, elite scientists, and bureaucrats. Inputs from these senior and elite stakeholders impacted problem formulation and the interventions we describe later in Incentives. Instructions and tools were handed off to the senior parties, to trickle down to field workers.

A vast majority of developers typecast field workers as data collectors. With the absence of direct interactions with field workers in most cases, the work of a field worker was primarily read through their dataset artefacts. Errors and aberrations in the dataset were reported to be a source of frustration for developers. Below, we describe the various ways in which developers conceived field workers—undergirding these conceptions is an equivalence assigned by developers to the roles of field workers and data collectors. Field workers were perceived to 'come in the way' of the worthwhile model development efforts of developers. Developers undertook various strategies to manage work practices of field workers, which we describe in the next section.

4.2 Conceptions of field workers as held by developers

The field worker as corrupt

Developers in our study frequently described field workers as morally corrupt, irresponsible, and fraudulent, when speaking of their work practices. The corruption conception was primarily ascribed by developers to field workers creating datasets with specific indicators like repeated values (*e.g.,* 100's of rows with same values) and expected values (*e.g.,* ideal blood pressures), *i.e.,* the datasets will filled and complete, but suspiciously ideal, a possible means of "*gaming the system*" (P65). Field workers were reported to falsify data entries for the purpose of passing manual or automated checks and acquiring remuneration. In some cases, developers first spoke of field worker corruption, and later acknowledged the tiresome effort involved in data collection. We remind that reader that our focus is on reporting developer views as-is, not concerning with these labels are accurate or valid for field workers. As we show in Interventions, many of these views are actioned upon in the form of interventions and processes introduced to the field workers by developers.

A common assumption among developers was that the field worker was deceiving and falsely passing off 'easy' work as 'real' work, abandoning the responsibilities they owed towards good data collection and model building. P17 working in Health AI explained how corruption occurred in data entry due to evaluation and incentive structures for data collection. *"[...]These surveyors get paid based on how many people they survey, so they will sit down underneath a tree and fill out form after form, because it's faster than to go and actually sit down and talk with people [...] The surveyors themselves are falsifying the data because of incentives." - P17, Health AI.* . In a few cases, developers labelled field workers as corrupt even if external constraints may have impacted data collection, *e.g.,* P26 spoke of "*fudged data*" coming from field workers entering average utility values if household residents were away, while reading electricity meters. Many developers spoke of data collection as a net good, expressing that the value of datasets was similarly valenced among all actors. (It is worth noting that prior studies demonstrate that field workers are often unaware of what and why they collect the datasets for [6, 41, 56].)

False data entries were reported to sabotage algorithm performance and robustness, due to the fact that models learned from incorrect entries that were not representative of the real world[3]. The concern for many developers was the possible occurrence of false positives and negatives due to false data, which can have enormous repercussions in low-resource contexts. Incorrect predictions in high-stakes domains were reported to be in the form of human, animal, and economic costs, *e.g.,* P55 working on healthcare models spoke of how false positives in Tuberculosis could wrongly route limited human resources in the backdrop of poor road and hospital infrastructure, and missing daily wages. False negatives were also reported to bear costs, as missed predictions were reported to sometimes mean life or death choices, *e.g.,* in time-sensitive, high-stakes predictions like advanced disease progression. A few developers spoke of 'corrupt' field workers as cunning and exploitative, sabotaging the model outcomes through wrong data. As P22, building a healthcare model spoke of ASHA (community health) workers, "*When I launched this product the first time [...] suddenly I was not getting fetal information for some pregnant woman. My doctor was telling me why this fetal information is missing here, maybe the baby was aborted. We rushed to the tribal village and interviewed the woman. The woman said, 'no health worker reached my home. Without coming here, how were they able to enter the data digitally'? Later we found out what this ASHA worker was doing with colleagues, collecting their data representing the pregnant mother. Then we understood that ASHA workers are very 'smart'*".

Since corruption among field workers was reported to manifest as ideal data values, this class was relatively straightforward to

[3]Fictitious data leading to high costs has been observed elsewhere *e.g.,* tens of thousands of fabricated phone numbers and addresses were found in COVID-19 tests during the Khumb Mela in 2021, which is widely believed to have led to a Delta surge [94].

detect, and hence identify offending field workers. Eyeballing, analysis tools, and AI models were used to detect these data entries. The value assignment of corruption to field workers was sometimes used as a rationale to build new models that automated out field workers, *e.g.,* building a computer vision AI model that is mandatory for a community health worker for use in visual health measurements, previously her job, motivated by the suspicion that the worker's manual data entries are fake. In some cases, field workers were required to provide cross-verification and documentation, *e.g.,* physical presence at the location, audio capture cross-verification, and observation notes. We discuss these strategies in greater detail in Interventions.

The field worker as lazy
Some developers described field workers as incompetent, stubborn, and sluggish by gathering data without sufficient care. Incompatible workflows, poor technical literacy, and obstacles in the data capture environment were cited as reasons for attributing inconvenience to field workers. Data collectors of datasets with missing values, flawed entries, and irregular entries were viewed as lazy. For example, image datasets often require ideal conditions during data capture, because models cannot typically handle high noise levels in the training datasets. In P59's case, working on agriculture and healthcare, the mandatory AI/ML requirement of white backgrounds in photographs captured in hospital or farm environments was physically impossible, due to the non-availability of white papers in farms, as well as the heterogeneous aesthetics in Indian settings, often painted with bright background colours. Some developers reported how informal practices of field workers were in collision with the AI/ML data entry requirements that were rather rigid in periodicity, location, and documentation-heavy.

While several developers acknowledged the difficulties faced by field workers in capturing the myriad fields of a dataset while attending to primary responsibilities, a few switched to pre-processing to deal with data capture difficulties (described in Interventions). Field workers were generally seen as cumbersome problems to be dealt with in AI/ML model development. Take the case of P45, building a model for crop yield detection, where farmers were required to collect 140 columns for every data entry row on a spreadsheet, yielding thousands of such data entries—done on a pro-bono basis, at the end of the day. Compliant farmers who entered well-labelled values were viewed as 'good' and 'professional' (P45), versus 'challenging' farmers who missed, delayed, or entered imperfect entries.

Field workers were seen as introducing human error in datasets, via workflows conflicts, forgetfulness, and non-diligence. However, the errors were often seen as costly and intentional aberrations, rather than mistakes. P69, building a health model, described how they perceived some frontline health workers to be lazy, "*Sometimes health workers come with data that might look suspicious. For example, once they observed all BP measurements as 80/120. So the NGO tried to investigate why. It could be because the health field worker was lazy or they did not have a BP machine and so they entered the typical values.*" The developer further described how public health doctors were later mandated to give feedback on patient data–a perceived productivity improvement over the 'slacking off' prior to their AI/ML model, "*90% of doctors never do the work, they don't even go to the PHC (clinic). Now we are making them do the work.*"

Some developers pointed to a lack of care on part of the field worker when entering data fields impacted by contextual peculiarities, such as changing phone numbers and shared devices (a phenomenon common to many regions of the Global South, as seen in [71]). P65 described the traditional norms, such as shared devices, as being inconvenient, "*5 people might use the same phone. People change SIM cards. How do you know it is the same person? Ownership of phones is important. This is traditional mindset that is harder to change.*" Field workers were also reported to enter their own phone numbers and addresses for requirements in data entry, when unable to or uninterested in comprehensively speaking with clients. Such entries were reported to be problematic for model training. In a few cases, practitioners themselves resorted to data collection in the field, as field workers were not compliant and did not see the value of models; thus, becoming a major inconvenience to developers. As P32 described, "*We have to go collect the data, because most of the farmers do not see the importance. They don't know that we are trying to help them.*" We underscore that P32's connotation was that their model carried tangible value for the field worker's community. However, the field worker was not seeing this value—a perception shared by some developers in our study.

The field worker as non-compliant
Some developers spoke of field workers as non-compliant actors, whose workflows and practices conflicted with the precision and accuracy required to build useful datasets. Developers reflected a sensory and mechanistic imagination of measuring the real world through recording devices, *e.g.,* capturing images, recorded sounds, tracked location, or clicked buttons. AI/ML data entry required periodic, standardized, and precise entries, such as capturing photographs of crops with ideal contrast levels and against clean backgrounds, or tracking exact river contours with GPS coordinates. 'Objective' dataset collection requirements were reported to be often in conflict with field worker workflows that were more informal, contingent, and situational in nature. For example, P27 reported how forest guards forgot to reset their GPS every 5 minutes; instead, by recording every one hour, the dataset became inaccurate and unfit for training their model. Developers that conceived of field workers as non-compliant often attributed distortions in data quality and accuracy, such as blurry images or noisy sound captures, to field workers.

The majority of developers appeared to understand the complex nature of humans using recording devices only well into the AI project development, and often by accident. Developers attributed the delayed insight to the fact that data collection was mediated through partners and middle-men. They also attributed it to the fact that these partners and middle-men typically had no prior interaction with field workers. P44 described how they learned about the influence of a medical machine operator on dataset quality only at an advanced stage of model development, pointing to the importance of data quality training, "*When we collected the data, we didn't have a lot of background on how the data was collected. We realized that the quality of the image at the O center was much better. Obviously it was a trained staff. It was a massive revelation for us, how an operator can influence this entire screening effort. If the nurse/operator is not trained, dataset suffers.*" Another example comes from P21, who discovered that 25-30% of their fetal cardio topography recordings from doctors were less than ten minutes long. P21 later worked with doctors to explain the value of the expected procedure to record data.

The projects falling within this category sought to typically enroll field workers to collect and label datasets for models that eventually aimed to diminish or scale their roles. The data collection requirements in this category entailed new and additional workflows to collect the data, *e.g.,* scan identity documents, traverse a forest to collect GPS data, and other tasks seemingly irrelevant to the field worker's primary objectives. The datasets discussed here were reported to be the hardest to fudge because they required hard recording evidence, *i.e.,* either there is a location on the map, or there is not. As P67 building a GIS model put it, "*Fudging GIS data is very difficult. In surveys it is easy to fake data. But here, unless you are physically at the space to collect the data, you cannot fake it.*" A few developers discussed how some of the field workers were concerned about potential job losses due to the models being built; in a couple of instances, field workers had organized protests to register their voices against the AI models. Developers spoke in our interviews of how these protests were a 'misunderstanding', ''personally driven', and 'politically motivated', indicating that workers attempted to sabotage their model building efforts.

The field worker as the dataset itself
When human work was not recognised in dataset collection, developers spoke of labour as the dataset itself, in the form of outcomes and outputs. In this category, complete, labelled datasets were the starting point of developer world view of AI/ML pipelines, *i.e.,* developers did not even speak of the initial labour of the field worker in collecting the dataset. In contrast, the remaining interviews show that collecting and building datasets was a monumental task involving field partners, infrastructure setup, special software, and regulatory requirements. Some developers did not concern themselves with dataset collection because it was deemed 'non-technical', outsourcing instead to NGOs, field partners and program managers. P67's response, despite their role as a program manager in-charge of dataset collection, is telling of how engaging with field workers was considered outside of the scope of AI/ML, "*My work is limited to technical challenges. I have never stopped to talk to the human side. To wonder how this (data collection) has made them feel.*". Many AI developers were far more comfortable speaking about missing, null, or inaccurate values over the human aspects of the data collection, even in projects with deep layers of human-mediated data collection. For example, P13 used the language of data discrepancies rather than human labour, "*It had a lot of missing data as well. There are some data points for which the dates don't match.*"

4.3 Interventions to manage data quality of field workers

In the final part of our results, we describe the various strategies reported by developers to improve the quality and quantity of data entries from field workers. The majority of interventions aimed to control or bypass field workers. In a handful of cases, interventions were empathetic towards the agency of the field workers, which we also report below. Developers reported that the management techniques, including monitoring, helped improve the data quality in many cases (albeit by possibly introducing an environment of fear, coercion, and extraction).

Surveillance
Developers outlined various monitoring approaches aimed at improving data entries in the field. field workers were reported to be continuously monitoring field workers via their personally identifiable data entries, at times, in combination with other sources like audio, location, and video snooping. Computational metaphors like sorting, assigning scores, stack ranking, and penalties were used in some interviews. For example, P18 spoke of assigning a probabilistic data quality score for live entries of field workers, "*some kind of a design of a data quality score, which is looking at the data collected by one worker. They'll assess if this data is good or bad quality on a score, based on 0-1*". In a few cases, poor performers were identified by comparing with data entries of other field workers. Non-compliant or non-performing field workers were notified directly, via motivational messages or reminders (we describe this technique in detail in Incentives). Supervisors and field officers were also alerted to take action on field workers, in some cases.

More complex and sophisticated forms of monitoring were reported in a few cases. P17 reported looking into a service that collected audio recordings of field workers "*while this person is giving this interview and then do some deep learning techniques to connect the audio quality with survey results to rate the surveyor. How clear are their questions? How much are they understanding the results? Make connections between survey fatigue and inaccuracies and consistency of the surveyor.*" Whether field workers were aware of these interventions being deployed on them was variable. P18 described how monitoring field workers improved the data quality, "*On a high level I can tell you that there are some known good healthcare field workers, because they're monitored, they're not completely on their own. We can treat healthcare data from those field workers as good.*" In other cases, field workers did not appear to be aware, or the developer was far removed from the field worker and was not aware of consent procedures. Prior work shows that surveillance affects employee well-being, productivity and motivation [5]; indeed, our findings underscore the shifting of risks of and costs from employers to field workers [69] and creating orthogonal, additional labour for the same or no pay.

Automation
Retrospective learnings and field insight led to pre-processing and technical fixes in some projects. Many developers reported making adjustments and fixes to data collection forms after receiving feedback from field partners, *e.g.,* making some fields mandatory, providing multiple choice options, decreasing the required radius of travel, and so on. In a few cases, pre-processing algorithms were introduced to reduce the burden of perfect audio or image capture by the field worker. In P59's case, a fix was written to process photos of pests from farms, dealing with the issue of non-availability of paper in farming areas. Other fixes included algorithms to deal with noise, blurriness, and duplications, at the time of data entry itself. Developers also described introducing techniques to seamlessly upload data entries over low memory, poor connectivity, and poor resolution cameras in devices used by field workers. For example, P23, 24, and 25 reported a fix they wrote to deal with the issue of variations in lighting and camera quality impacting the model predictions.

Despite the above examples easing the workload and usability among field workers, automation also had a dual use of bypassing field workers and their agency. A few developers revealed how automated gathering of ground truth was increasingly used as a proxy that eliminated the field worker's human entries. Automation was sometimes used to override or bypass the field worker, *e.g.,* in the case where a developer created a wearable for automatic data collection to be directly worn by pregnant mothers, and the field worker performed the role of a network transporting the data from patient to their tablet. The intervention was reported to be created to manage fudged data, for field workers to be "*really visiting their (patient) homes*".

Incentives and gamification

In some cases, intrinsic and extrinsic incentives were deployed to field workers to motivate better performance and recognize good performance. Many developers turned to tactics like points, leaderboards, and positive or congratulatory messages and calls to improve field worker goodwill and motivation. In a few cases, developers reported making payments to field workers for data collection. Some spoke of wanting to make payments to field workers "*In the future*" (P26). Developers experimented with various payout amounts and frequencies, aiming to balance quality of data entries with field worker satisfaction. P14 reported moving from a monthly incentive to a weekly format for incentive payments, pointing to how cash flows are limited among drivers. P51 described how incentives led to skews in distribution, "*they stopped farming and started focusing on the data. So then they brought in thousands of images in a day, and it skewed the region*" In a few cases, developers were successfully able to tie the project needs to the resources of the institutions or state bodies, which greatly eased their reach and access. For example, P22 was able to peg the health worker performance to a state body, leading to a bonus payment, promotion, or recognition by the district officers in monthly meetings. P37 articulated the value of partnering with similarly motivated institutions. "*It's important that the work is the goal of the hospital too. We need a hands-on research person like him because we are not always there.*" Nonetheless, incentives were largely automated (*e.g.,* motivational SMS) or tied to senior field partners, with the exception of a couple of developers that worked with field workers directly to determine amounts and make payments. A few developers discussed the goodwill of communities to help out the AI developers, both as a lucky outcome ("*inherent advantage of coming from [an elite institution] got the goodwill of people working with us. Even the patients gave consent, we did not give any incentive to them*" (P21) and as a manipulative tactic used by other developers, *e.g.,* P67 spoke of how college students, in their desire to gain street credibility and career experience, were motivated to contribute towards open data collection efforts pro-bono, even with only partial understanding of the uses of the data they collected.

Direct engagement

In a minority set of cases, developers directly engaged with field workers in their dataset development. By and large, even among those following engagement strategies, we did not encounter instances of field workers being spoken of as experts, except when reporting on interactions with professionals such as agricultural scientists or oncologists. Some developers bemoaned that they should have reached out to 'untapped' expertise in field workers, in retrospect, but their perceived technical limits of their jobs kept them from establishing contact. P54's view is reflective of this stance, "*I'm not sure what people who are collecting the data expect this data to be, [know] why it's important to flag something, cannot put arbitrary numbers. I'm not sure if that connection happens, if they know how fruitful their work is, not just a checklist, but will go into a learning framework, predict something really important for patients.*" At the same time, a few developers did engage with field workers at various stages. P50 reported how until speaking with a field worker in the region, "*and they told us not to rely on that dataset, we were predicting all kinds of weird species*".

A few developers conducted data literacy camps and educated field workers on the value of the models. For example, P14, building an air quality prediction model that engaged informal transport drivers (boda bodas) to collect air pollution readings. P14 organised a workshop with medical doctors on air quality and health for the drivers to attend, which helped build trust. P14 reported, "*The riders felt more important, asked questions and got to know more about the subject. They feel like they're experts amongst the community. They felt empowered.*" After observing gynaecologists and explaining the value of longer ultrasound recordings, P21 determined that field visits led to improvements in data quality. A few developers discussed how communicating the impact of data quality collected by field workers on technical outcomes like accuracy and performance was challenging. As P51 reported, "*If you think of impact in AI, it would be I created a new algorithm, published a paper, is it exciting technically. It improved a body of knowledge. But for the small farmer there, it does not prove impact if I increase the accuracy 80-85%.*" Explaining potential application outcomes was considered more comprehensible for field workers, *e.g.,* explaining how models can have an impact on disposable income, better health, or kids staying in school, but showing those impacts was reported to be challenging, especially for models still in development. The low-resource and high-stakes nature of models in our study meant that the tangible impact after deployment took a few years to manifest. A few developers made visits to sites of field workers, to demonstrate the intangible value of their data collection, such as P7 and their team of medical students who visited nurses doing data collection, "*most trivial thing is to just be a visitor and see if a sensor was actually off, like an oxygen clip on the finger. Just paying a visit, letting them know that we care about these things. Showing these signals to the nurses on the interface [and demonstrating the importance].*"

Developers reported engaging with field workers much more when the downstream beneficiaries of their models were end-users, *e.g.,* training nurses to collect patient data in hospitals. In a few cases, the users themselves were data contributors, *e.g.,* for maternal wearables collecting pregnant women's data in trial settings. User consent, education, and compliance was seriously regarded, citing institutional and regulatory requirements. For example, P10, P11 and P12 who were building a wearable sensor for a health model, planned training sessions for every data collector on usability, troubleshooting, and proper positions.

5 DISCUSSION

We find that in their search for building reliable datasets, AI developers reduced field workers employed to provide socio-economic benefit (*e.g.,* a community health worker to track COVID-19 incidence and administer vaccines), to a data collector. The field workers were tapped for additional and often monotonous labour of dataset collection such as quantising, counting, sorting, tagging, clicking, and recording various phenomena in the real world. Field workers was deskilled and they were recruited as menial data collectors with no expertise. As People's Archive of Rural India reported in 2020, overburdened ASHA (community health) workers in India are tasked with a mind-boggling array of 720 tasks, including data collection tasks [6]. Models emulated and improved over expertise of field workers, but the expertise—which takes many years to build—was only seen as an instrumental tool for collecting datasets, and largely excluded from model development.

Developers reported that many field workers ignored, fudged, and sabotaged data collection, producing data that was indiscernible or unsatisfactory for models. Correspondingly, developers attributed the problem of poor data sets to poor work practices of field workers. Unfortunately, these refractory images of workers were commonly held among developers, even though most developers had no firsthand understanding of worker practices or contact with workers. Instead, worker outputs were read through their dataset artifacts. Through various interventions, developers attempted to influence what they perceived as incompetent, lazy, corrupt practices of data collectors to improve data quality. Indeed, the data collection issues like fudging and sabotage are commonplace in working with field workers in low-resource areas, but it is possible to interpret these acts as worker resistance to their exploitative conditions of work, and not as incompetance and laziness on part of workers (as seen by developers)—modifying James C. Scott's memorable phrase, '*weapons of the data collector*' [75]. We call for more research on field workers' perceptions and their experiences in being drafted into building AI systems to understand this problem space better.

Below, we discuss considerations and open questions for recognizing field workers as domain experts, and for practitioners to include field workers in problem formulation and building datasets and models, and propoe structural changes to recognise domain expertise in the field of AI/ML.

5.1 What counts as expertise? Whose expertise counts?

Developers in our study animated their model building with visions of producing new forms of scalable expertise to fix socio-economic problems (similar to high-modernist visions in Scott's account [74]). Field workers were needed solely as computational resources, whereas the elite status of AI system building was restricted to the developers, leaders of partner organisations, celebrity scientists, and bureaucrats, and the machine intelligence itself (see Ivan Illich for more on experts and gate-keeping [37]). A spectre of expertise paucity was often maintained, which perpetuated the demand for AI models. For example, farmers were viewed as recording devices, whereas heads of agricultural organisations were viewed as partners occupying a certain cultural legitimacy and authority. Discounting expertise of "data collectors" is not limited to our study, being well-documented throughout the history of computing, *e.g.*, in how women technologists [34], business process outsourcing workers [57], data workers [52, 89], and crowd workers [21, 39] have been perceived as low-cost, unskilled workers, despite performing valuable services. It is also the case that there has not been much thought given to who is recruited as a data collector. Field workers were automatically drafted into the surplus labour of data collection when heads of organisations signed contracts and started partnerships with developers. Finally, in contrast to their typecasting of field workers as data collectors, developers spoke reverentially of end-users, taking seriously user design considerations such as usability, languages, transparency, and explainability (owing in part to policy and market value of user loyalty).

We make the case that the mastery and tacit knowledge of communities and contexts held by field workers need to be accurately located as domain expertise by AI developers. The domain expertise can serve as an essential partnership throughout the AI pipeline. Such collaborative partnerships were exceptional in our study, underscoring why we need more sustained and structural shifts in the field of AI/ML to make recognition for domain expertise of field workers a part and parcel of 'doing AI in low-resource areas'. Recognizing the expertise of field workers is not merely about ethics; it successfully sets up the AI engineering fundamentals of getting consistently good quality data, timely feedback on deployments, and confidence in building solutions. Recognizing expertise further demystifies what AI is, who is involved in doing AI, and whose labour counts as AI development. A crucial step in this direction is for developers to engage directly with field workers, instead of interfacing with organisation leaders and hearing secondhand narratives about workers. As Lucy Suchman points out, "*work has a tendency to disappear at a distance, such that the further removed we are from the work of others, the more simplified, often stereotyped, our view of their work becomes*" [79].

Field workers could be involved right from the problem formulation stage, to accurately characterise the problem to be solved, making assumptions about models, and identifying outcomes salient for predictions and classification tasks together. Partnership at this stage can be enormously beneficial in setting the foundation for the project, aiding more accurate and relevant problem formulation, *e.g.,* in the case of P27, who formulated a more impactful poaching prediction problem with wildlife experts. Dataset parameters, such as data fields, survey volume, and answer ranges, can be jointly identified, based on feature requirements and field norms. Observing the workflows and practices of field workers can help contextually fit dataset tasks. For example, P14 visited boda boda (motorcycle taxi) drivers and rode along as they collected air quality data, identifying several obstacles to dataset collection and the need for better incentives. Although a few practitioners in our study recognized the untapped expertise of field workers, we need a sustained effort to broaden such collaborations.

A general maxim in current AI praxis is that the more data we have, the better performing the model will be. However, defining dataset tasks without an understanding of the everyday realities of field workers can miss out on valuable indicators that could be collected. Developers should inform field workers about the downstream use of these datasets, the types of use cases and users imagined, and the impact aspired for, *e.g.,* in the case of P64, who visited overworked hospital nurses and assured them of the importance of their work. Field workers may need dataset collection training to understand question goals, troubleshoot, handle response ambiguity, and so on. Developers could benefit from serving as instructors of training programs, and receive feedback on their data procedures.

During model development, developers should involve field workers in identifying modelling goals, features, and measurement criteria. Field workers can enable 'online learning' in low-resource settings, by creating feedback loops between model outcomes and the real-world, and help evaluate whether pilot results are accurate and valid on the ground. Indeed, partnership in data and model building can help identify and alleviate algorithmic bias, data cascades, and unintended outcomes that typically remain latent and opaque for long time periods [42, 46, 72]. Understandably what drives metrics currently is driven by concerns of machine-beats-human efficiency, cost, and outperforming the industry standard, which often relies on the expertise of AI developers. We need to rely on the expertise of field workers to identify appropriate metrics to evaluate claims relevant to the affected communities and phenomena. Where regulations permit compensation for AI labour, pay needs to be fair and commensurate to contributions to AI systems.

5.2 When to intervene? How to intervene?

We found that management of poor data collection practices generated a whole class of algorithmic fixes for AI/ML developers. Pre-processing algorithms and regular expression checks in form filling allowed developers to handle heterogeneity in environments and field workers. These techniques were arguably the most mutually beneficial by reducing human drudgery and improving data quality. At the same time, entire models were developed to predict data outliers and corresponding field workers who appeared to violate certain heuristics of data entry, such as those entering predictable or repeated values. For example, newborn birth weight entries by field workers can be tested by a constraints check on inputted values, automatically flagging non-compliant field workers. This information can be asymmetrically shared with higher-up supervisors or used in sending gamified messages to the field workers to do better at data collection, thus worsening oppression and enabling extraction. Currently, what gets automated depends solely on the preoccupation of AI developers' concerns and personal values, which can be limiting and even dangerous for domain experts and communities. For example, most developers in our study did not consider relieving field workers' concerns a legitimate need for AI; instead their primary objective was to introduce a model where there was a good quality dataset and a rhetorically defensible socio-economic cause. However, many of these AI interventions did not appear to fundamentally address the root problem and instead appeared to create new forms of displacement and extraction—we invite more empirical evaluations here. Instead of understanding the difficulties encountered in the data collection processes by consulting field workers, AI developers sought to develop behavioral nudges and interventions to discipline and control field workers.

Behavioural tricks like gamification and monitoring can introduce fear and resentment among workers [50]. Recent emergence of interdisciplinary research across AI and behavioural science aims at influencing human behaviour through nudges, rewards, gamification, personalisation, and segmentation [64]. In this way, algorithms are seen as contested instruments of control that carry specific ideological preferences [92], which are often created and implemented based on the interests of powerful actors [43, 87]. If developers consider the task of data collection as a challenging and important problem to solve, rather than an impediment that blocks the model building stage, it would result in an attitude change in how they see the work of field workers. Rather than trying to create techniques to gamify, monitor and somehow persuade field workers to do work, if AI developers actually considered data collection as legitimate work (data work is broadly under-recognized in AI, see [72]), possibilities to figure out which aspect of the work can be meaningfully automated can be sought out.

Instead of motivating an overworked health worker to do more work for dataset collection, one might ask how to help them achieve their goals better: prioritise their numerous visits, better capture of their in-situ knowledge, allocate limited medical resources better, and build visibility into their contributions. Indeed, this is one way to realize the shared goals of both the AI developer and field workers on human development in low-resource areas. As an example, Digital Green [27] recognised and tapped into the expertise of farmers and turned them into producers of expert videos, and found appropriate ways to scale up that knowledge. Farmers were motivated to learn from other farmers, as well as from other agricultural scientists, and a meaningful interaction with digital technology led to the possibility of skilling of their expertise.

The process of building AI models is power-laden with consequences that may undermine the AI enterprise. Consider this thought experiment, where the tables are turned and the field workers decide to monitor the work of AI developers. It would be an unthinkable scenario; apart from the power asymmetry, we implicitly value and trust the work done by the AI developer. Why should the current hegemonic arrangement of AI experts monitoring field workers makes sense? This is because, to borrow from James Ferguson, data collection in AI operates as an 'anti-politics machine' [23], making political decisions about what constitutes expertise and creating 'technical solutions to technical problems' through management interventions to influence field workers to collect good quality data.

5.3 How can domain expertise be made legible?

The untapped expertise of field workers we describe is not an exception or error; it is intricately tied to any AI development in low-resource settings [73]. In contrast, the standard AI/ML pipeline often begins with an existing or available dataset—an assumption shaped by institutional and resource endowments and prior histories of the West. We need to recognise that expertise is required in all stages of the pipeline, not just by the developer at the model development stage. In particular, we call for the need to make legible the labour of field workers in the early stages of the pipeline, the work that remains below the API fold. The field of AI/ML does not currently have language, standards, requirements, or measurements for discussing the recognition of domain expertise. These shifts cannot be left to the whims and values of individual research projects, and instead need broader changes and oversight in the field. Some shifts can come from releasing disclosures that go beyond fact sheets that are typically focused on a technical view of data quality, transparency, and bias [51]. We need to find ways to disclose the provenance, type, and amount of labour that went into building powerful AI systems, similar to impact statements that are now standard in AI conferences [11]. In addition to fair compensation, field worker's contribution to the algorithm can be made visible in many ways, including co-authorships, listing contributors publicly, joint releases, and introducing the field workers in high visibility events. Indeed, attribution raises questions of fairness, politics, and decoloniality, such as whose knowledge is considered attributable and whose is not [67].

It is imperative to expand the parameters of what gets measured and seen, in order to fully comprehend the effects of these AI systems. Developers talked to us about how they could create new expertise in algorithms that saves human lives, tiger lives, farm fields, money and so on, all of which they were keen to assess the impact of. Vision statements and impact metrics of projects in our study were shaped by prevailing economic impact assessments of AI systems; these assessments are coloured by conventional pipeline-thinking, where algorithms are dropped into social settings and their before-after effects are measured, *e.g.,* the number of lives saved by the algorithm [93]. Missing in these variables and constructs is the measurement of the hidden labour of field workers (among others, including data annotators and content moderators) in enabling the systems. As mentioned earlier, data in low-resource contexts is not a given; it comes into existence often via the labour of the field workers.An excellent development is the work done by the Indian Federation Of App-Based Transport Workers to resort to social media to raise awareness of rights of "gig workers" to demand fair credit. Thus, we make the case to focus not just on the narrow goals of model outcomes, but also on the effects on the humans who are all part of the AI enterprise. More research is needed to understand the longer-term implications of labour and expertise re-distribution in AI. Deployments should consider these systems as interventions that require the labour of field workers to come into existence and become sustainable. Ultimately, our direct ask to AI developers is to embody a reflexivity to their practice, what Philip Agre called a "critical technical practice" with "one foot planted in the craft work of design and the other foot planted in the reflexive work of critique" [2].

6 CONCLUSION

AI development apparatus in low-resource areas largely deskills and invisibilises the domain expertise of field workers. Field workers are indispensable contributors to collecting datasets for AI models in low-resource settings, but often were not recognized for their expertise by AI/ML developers. In our research study with 68 globally-distributed AI/ML developers building models for low-resource communities, we find that field workers, *e.g.,* forest guards and farmers, are treated as data workers in service of the machine. AI developers demonstrated a belief that 'bad data' came from poor data collection practices of field workers in the following ways: as corrupt, lazy, non-compliant, untapped expert and data itself. Correspondingly, AI developers often sought to to discipline and control the work practices of labourers in dataset collection. The field worker was viewed as a data worker, reducing their embedded domain expertise into counting and clicking for ML datasets, despite the labourers not being aware of why, what, or how their work was impacting AI/ML models, or being able to consent or object to the process. Thus, AI development in low-resource areas appears to deskill the domain expertise of field workers, despite models seeking to improve over their expertise in the form of datasets. Field workers appeared to resist these designs by falsifying or producing incomplete data. Our research results provide novel and crucial insight into how AI development practices function as a deskilling apparatus in low-resource contexts. We press upon the HCI community the need for questioning what expertise is, by partnering and recognising the local expertise of field workers, formulating better problems that improve worker agency rather than seeking to control, and to fairly account for field workers contributions to AI models.

ACKNOWLEDGMENTS

We are grateful to the various AI practitioners who took part in our study and generously shared their AI development experiences with us. Our thanks to Hannah Highfill, Diana Akrong, and Shivani Kapania for contributing to the research interviews.

REFERENCES

[1] Mark S Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work (CSCW)* 22, 4-6 (2013), 531–573.
[2] Philip E Agre. 1997. Lessons learned in trying to reform AI. *Social science, technical systems, and cooperative work: Beyond the Great Divide* 131 (1997).
[3] Ali Alkhatib and Michael Bernstein. 2019. Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[4] Paul Attewell. 1987. The deskilling controversy. *Work and occupations* 14, 3 (1987), 323–346.
[5] Kirstie Ball. 2010. Workplace surveillance: An overview. *Labor History* 51, 1 (2010), 87–106.
[6] Anubha Bhonsle and Pallavi Prasad. 2020. Counting cows, not rural health indicators. https://ruralindiaonline.org/articles/counting-cows-not-rural-health-indicators/. (Accessed on 08/02/2021).
[7] Harry Braverman. 1998. *Labor and monopoly capital: The degradation of work in the twentieth century*. NYU Press.
[8] John Seely Brown and Paul Duguid. 2017. *The social life of information: Updated, with a new preface*. Harvard Business Review Press.
[9] Michael Burawoy. 1990. *The politics of production*. Verso London.
[10] Jenna Burrell and Marion Fourcade. 2020. The Society of Algorithms. *Annual Review of Sociology* 47 (2020).
[11] Davide Castelvecchi. 2021. Prestigious AI meeting takes steps to improve ethics of research. *Nature* (2021), 12–13.
[12] Kuang Chen, Joseph M Hellerstein, and Tapan S Parikh. 2011. Data in the First Mile.. In *CIDR*. Citeseer, 203–206.
[13] Ed H Chi. 2017. Technical perspective: Humans and computers working together on hard tasks. *Commun. ACM* 60, 9 (2017), 92–92.
[14] Nicola Dell, Trevor Perrier, Neha Kumar, Mitchell Lee, Rachel Powers, and Gaetano Borriello. 2015. Digital Workflows in Global Development Organizations. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1659–1669.
[15] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, and Morgan Klaus Scheuerman. 2020. Bringing the people back in: Contesting benchmark machine learning datasets. *arXiv preprint arXiv:2007.07399* (2020).
[16] Tom Denton. 2016. Seeing like a statistical learning algorithm. https://inventingsituations.net/2016/07/10/seeing-like-a-statistical-learning-algorithm/
[17] Cory Doctorow. 2021. Pluralistic: 19 Aug 2021 – Pluralistic: Daily links from Cory Doctorow. https://pluralistic.net/2021/08/19/failure-cascades/#dirty-data. (Accessed on 09/01/2021).
[18] Jonathan Donner, Rikin Gandhi, Paul Javid, Indrani Medhi, Aishwarya Ratan, Kentaro Toyama, and Rajesh Veeraraghavan. 2008. Stages of design in technology for global development. *Computer* 41, 6 (2008), 34–41.
[19] Paul Duguid. 2012. *'The art of knowing': social and tacit dimensions of knowledge and the limits of the community of practice*. Brill Sense.

[20] Kate Ehrlich and N Sadat Shami. 2008. Searching for expertise. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. 1093–1096.

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

[22] Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data augmentation for low-resource neural machine translation. *arXiv preprint arXiv:1705.00440* (2017).

[23] James Ferguson. 1990. *The anti-politics machine:'development', depoliticization and bureaucratic power in Lesotho.* CUP Archive.

[24] Deborah Fitzgerald. 1993. Farmers deskilled: Hybrid corn and farmers' work. *Technology and Culture* 34, 2 (1993), 324–343.

[25] Carl Benedikt Frey. 2019. *The technology trap.* Princeton University Press.

[26] Carl Benedikt Frey and Michael A Osborne. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change* 114 (2017), 254–280.

[27] Rikin Gandhi, Rajesh Veeraraghavan, Kentaro Toyama, and Vanaja Ramprasad. 2007. Digital green: Participatory video for agricultural extension. In *2007 International conference on information and communication technologies and development*. IEEE, 1–10.

[28] Ishita Ghosh, Jay Chen, Joy Ming, and Azza Abouzied. 2015. The persistence of paper: a case study in microfinance from Ghana. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*. 1–10.

[29] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT press.

[30] Mark Graham and Jamie Woodcock. 2018. Towards a fairer platform economy: Introducing the fairwork foundation. *Alternate Routes* 29 (2018).

[31] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass.* Eamon Dolan Books.

[32] David Alan Grier. 2013. *When computers were human.* Princeton University Press.

[33] The World Bank Group. 2017. Trouble in the Making? The Future of Manufacturing-Led Development. http://www.worldbank.org/en/topic/competitiveness/publication/trouble-in-the-making-the-future-of-manufacturing-led-development.

[34] Mar Hicks. 2017. *Programmed inequality: How Britain discarded women technologists and lost its edge in computing.* MIT Press.

[35] J Hirschberg. 1998. Every time I fire a linguist, my performance goes up, and other myths of the statistical natural language processing revolution. Invited talk. In *Fifteenth National Conference on Artificial Intelligence (AAAI-98)*.

[36] Ursula Huws. 2014. *Labor in the global digital economy: The cybertariat comes of age.* NYU Press.

[37] Ivan Illich. 1975. *Medical nemesis.* Citeseer.

[38] Lilly Irani. 2015. The cultural work of microwork. *New Media & Society* 17, 5 (2015), 720–739.

[39] Lilly Irani. 2019. Justice for Data Janitors. In *Think in Public.* Columbia University Press, 23–40.

[40] Lilly C Irani and M Six Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 611–620.

[41] Azra Ismail and Neha Kumar. 2018. Engaging solidarity in data collection practices for community health. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[42] Nathan Kallus and Angela Zhou. 2018. Residual unfairness in fair machine learning from prejudiced data. In *International Conference on Machine Learning*. PMLR, 2439–2448.

[43] Katherine C Kellogg, Melissa A Valentine, and Angele Christin. 2020. Algorithms at work: The new contested terrain of control. *Academy of Management Annals* 14, 1 (2020), 366–410.

[44] Meghana Kshirsagar, Caleb Robinson, Siyu Yang, Shahrzad Gholami, Ivan Klyuzhin, Sumit Mukherjee, Md Nasir, Anthony Ortiz, Felipe Oviedo, Darren Tanner, et al. 2021. Becoming Good at AI for Good. *arXiv preprint arXiv:2104.11757* (2021).

[45] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge university press.

[46] Lydia T Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. 2018. Delayed impact of fair machine learning. In *International Conference on Machine Learning*. PMLR, 3150–3158.

[47] David Lyon. 2013. *The electronic eye: the rise of surveillance society-computers and social control in context.* John Wiley & Sons.

[48] James Manyika. 2017. What technology can do for Africa. https://www.mckinsey.com/featured-insights/employment-and-growth/technology-jobs-and-the-future-of-work.

[49] David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 224–235.

[50] Alexandra Mateescu and Aiga Nguyen. 2019. "Workplace monitoring and surveillance".

[51] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[52] Naja Holten Møller, Claus Bossen, Kathleen H Pine, Trine Rask Nielsen, and Gina Neff. 2020. Who does the work of data? *Interactions* 27, 3 (2020), 52–55.

[53] Laura Nader. 1972. Up the anthropologist: Perspectives gained from studying up. (1972).

[54] Bonnie A Nardi and Yrjö Engeström. 1999. A web on the wind: The structure of invisible work. *Computer supported cooperative work* 8, 1-2 (1999), 1–8.

[55] Matthew J O'Brien, Allison P Squires, Rebecca A Bixby, and Steven C Larson. 2009. Role development of community health workers: an examination of selection and training processes in the intervention literature. *American journal of preventive medicine* 37, 6 (2009), S262–S269.

[56] Chinasa T Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It cannot do all of my work": Community Health Worker Perceptions of AI-Enabled Mobile Health Applications in Rural India. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–20.

[57] Jacki O'neill and David Martin. 2013. Relationship-based Business Process Crowdsourcing?. In *IFIP Conference on Human-Computer Interaction*. Springer, 429–446.

[58] Joyojeet Pal, Anjuli Dasika, Ahmad Hasan, Jackie Wolf, Nick Reid, Vaishnav Kameswaran, Purva Yardi, Allyson Mackay, Abram Wagner, Bhramar Mukherjee, et al. 2017. Changing data practices for community health workers: Introducing digital data collection in West Bengal, India. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*. 1–12.

[59] Lawrence A Palinkas, Sarah M Horwitz, Carla A Green, Jennifer P Wisdom, Naihua Duan, and Kimberly Hoagwood. 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health and mental health services research* 42, 5 (2015), 533–544.

[60] Tapan S Parikh. 2009. Engineering rural development. *Commun. ACM* 52, 1 (2009), 54–63.

[61] Samir Passi and Steven Jackson. 2017. Data vision: Learning to see through algorithmic abstraction. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 2436–2447.

[62] Andrew Perrault, Fei Fang, Arunesh Sinha, and Milind Tambe. 2019. AI for social impact: Learning and planning in the data-to-deployment pipeline. *arXiv preprint arXiv:2001.00088* (2019).

[63] Fahad Pervaiz, Aditya Vashistha, and Richard Anderson. 2019. Examining the challenges in development data pipeline. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 13–21.

[64] Ganna Pogrebna. 2019. Behavioural data science. https://www.turing.ac.uk/research/research-programmes/finance-and-economics/behavioural-data-science

[65] Ghana Pooneh. March 2018. The Future of Work. https://www.wired.com/brandlab/2018/03/the-future-of-work.

[66] Philip L Powell, Jonathan H Klein, and NAD Connell. 1993. Experts and expertise: the social context of expertise. In *Proceedings of the 1993 conference on Computer personnel research*. 362–368.

[67] Mariolga Reyes Cruz. 2008. What if I just cite Graciela? Working toward decolonizing knowledge through a critical ethnography. *Qualitative Inquiry* 14, 4 (2008), 651–658.

[68] Ruth G Rinard. 1996. Technology, deskilling, and nurses: the impact of the technologically changing environment. *Advances in Nursing Science* 18, 4 (1996), 60–69.

[69] A Rosenblat, T Kneese, and B Danah. 2014. Workplace surveillance data & society. *Thousand Oaks, CA* (2014).

[70] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, and Kristy Milland. 2015. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1621–1630.

[71] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. 2018. " Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*. 127–142.

[72] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 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*. 1–15. https://doi.org/10.1145/3411764.3445518

[73] Nithya Sambasivan and Thomas Smyth. 2010. The human infrastructure of ICTD. In *Proceedings of the 4th ACM/IEEE international conference on information and communication technologies and development*. 1–9.

[74] James C Scott. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Yale University Press.

[75] James C Scott. 2008. *Weapons of the Weak.* Yale University Press.

[76] Zheyuan Ryan Shi, Claire Wang, and Fei Fang. 2020. Artificial Intelligence for Social Good: A Survey. arXiv:2001.01818 [cs.CY]

[77] Yaron Singer and Manas Mittal. 2013. Pricing mechanisms for crowdsourcing markets. In *Proceedings of the 22nd international conference on World Wide Web*. 1157–1166.

[78] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)* 8, 1 (1999), 9–30.

[79] Lucy Suchman. 1995. Making work visible. *Commun. ACM* 38, 9 (1995), 56–64.

[80] Lucy Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.

[81] Sharifa Sultana, Md Mobaydul Haque Mozumder, and Syed Ishtiaque Ahmed. 2021. Chasing Luck: Data-driven Prediction, Faith, Hunch, and Cultural Norms in Rural Betting Practices. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[82] Astra Taylor. 2018. The automation charade. *Logic Magazine* 5, 1 (2018).

[83] Divy Thakkar, Azra Ismail, Alex Hanna, Pratyush Kumar, Nithya Sambasivan, and Neha Kumar. 2022. When is ML data good?: Valuing in Public Health Datafication. In *proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

[84] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.

[85] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle CM Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, et al. 2020. AI for social good: unlocking the opportunity for positive impact. *Nature Communications* 11, 1 (2020), 1–6.

[86] Rajesh Veeraraghavan. 2013. Dealing with the digital panopticon: the use and subversion of ICT in an Indian bureaucracy. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers-Volume 1*. 248–255.

[87] Rajesh Veeraraghavan. 2021. Cat and Mouse Game: Patching Bureaucratic Work Relations by Patching Technologies. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[88] Michael Wallace and Arne L Kalleberg. 1982. Industrial transformation and the decline of craft: The decomposition of skill in the printing industry, 1931-1978. *American Sociological Review* (1982), 307–324.

[89] Ding Wang, 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*.

[90] Etienne Wenger. 2011. Communities of practice: A brief introduction. (2011).

[91] Sarah Myers West, Meredith Whittaker, and Kate Crawford. 2019. Discriminating systems: Gender, race and power in AI. *AI now Institute* (2019), 1–33.

[92] Langdon Winner. 1980. Do artifacts have politics? *Daedalus* (1980), 121–136.

[93] Justus Wolff, Josch Pauling, Andreas Keck, and Jan Baumbach. 2020. The economic impact of artificial intelligence in health care: systematic review. *Journal of medical Internet research* 22, 2 (2020), e16866.

[94] Sameer Yasir. 2021. Fake Coronavirus Tests May Have Helped Fuel India Outbreaks - The New York Times. https://www.nytimes.com/2021/06/15/world/india-fake-coronavirus-covid-tests.html. (Accessed on 08/11/2021).

[95] Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon. 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In *Thirty-First AAAI conference on artificial intelligence*.

[96] Shoshana Zuboff et al. 1988. In the age of the smart machine. (1988).