Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices

Milagros Miceli
Technische Universität Berlin
m.miceli@tu-berlin.de

Tianling Yang
Technische Universität Berlin
tianling.yang@tu-berlin.de

Laurens Naudts
Centre for IT & IP Law (CiTiP), KU Leuven
laurens.naudts@kuleuven.be

Martin Schuessler
Technische Universität Berlin
schuessler@tu-berlin.de

Diana Serbanescu
Technische Universität Berlin
diana-alina.serbanescu@tu-berlin.de

Alex Hanna
Google Research
alexhanna@google.com

ABSTRACT

In industrial computer vision, discretionary decisions surrounding the production of image training data remain widely undocumented. Recent research taking issue with such opacity has proposed standardized processes for dataset documentation. In this paper, we expand this space of inquiry through fieldwork at two data processing companies and thirty interviews with data workers and computer vision practitioners. We identify four key issues that hinder the documentation of image datasets and the effective retrieval of production contexts. Finally, we propose reflexivity, understood as a collective consideration of social and intellectual factors that lead to praxis, as a necessary precondition for documentation. Reflexive documentation can help to expose the contexts, relations, routines, and power structures that shape data.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; • **Social and professional topics → Quality assurance**; *Computing industry*; • **Computing methodologies** → Computer vision problems.

KEYWORDS

datasheets for datasets, dataset documentation, documentation obstacles, deflexivity, transparency, accountability, audits, machine learning, computer vision datasets, data annotation, image data, data creation, work place ethnography, image labeling, power

ACM Reference Format:
Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442188.3445880

FAccT '21, March 3–10, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8309-7/21/03.
https://doi.org/10.1145/3442188.3445880

1 INTRODUCTION

Since the rise of deep learning and convolution neural nets, the field of computer vision has demonstrated some of the most impressive results in machine learning [41]. Reaching a new high in popularity, computer vision models are used in a broad range of applications, penetrating ever more aspects of daily life. Creating datasets for computer vision is not straightforward. Work practices involved in gathering, annotating, and cleaning image data comprise subjective choices and discretionary decision-making [35, 39, 40]. Such decisions range from the framing of real-world questions as computational problems [5, 38] to the establishment of taxonomies to label images [32]. Data is also "the product of unequal social relations" [19] that are present among data workers as well as in the relationship between those whose data is collected and those who make use of data for research and/or profit. The opacity of industrial practices regarding computer vision datasets is a significant threat to ethical data work and intelligible systems [49].

Recent research has proposed implementing structured disclosure documents to accompany machine learning datasets [4, 22, 23, 27]. Despite their good intentions, those efforts fail to effectively reflect power dynamics and their effects on data [19, 32]. For instance, Gebru et al. [22] propose that datasheets include the question "does the dataset identify any subpopulations?" [22] e.g. by race, age, or gender. This way of documenting dataset composition is helpful. However, we argue that disclosing if a dataset includes racial categories does not speak to the problem of such categories' reductiveness, nor makes the assumptions behind race classifications embedded in datasets explicit. In the same way, asking "who created this dataset?" [22] and "who was involved in the data collection process (...) and how were they compensated?" [22] remains insufficient to interrogate hierarchies in industrial settings and their effects on data [32] . Reflecting on interests, preconceptions, and power encoded in training data [16, 19, 46] is essential for addressing many of the ethical concerns surrounding computer vision products.

In this paper, we lay our focus at the intersection of manual data processing and computer vision engineering. We investigate how work practices involved in the production of computer vision datasets can be made explicit in documentation. Although data processing can cover a variety of activities, we refer to companies where human workers collect, segment, and label image training data. Data processing companies of this kind provide data services at the request of computer vision companies (hereinafter "requesters") that wish to outsource parts of dataset production. Work between service providers and requesters requires strong coordination efforts as it comprises many actors and iterations [32]. Collaboration is informed by negotiation over the meanings that are ascribed to images [16]. In this context, not all actors hold equal power to shape datasets: Data processing companies generally collect and interpret data according to categories instructed by requesters, and workers often trust the judgment of their managers in case of doubt or disagreement. Power to decide how images are to be interpreted, classified, and labeled depends on the position occupied by each actor [32]. These dynamics have a crucial effect on the datasets that train commercial computer vision products. Making them explicit in documentation can help better understand models' behavior and uncover broader ethical issues.

We base our investigation on fieldwork at two data processing companies, and several expert interviews with data quality analysts and computer vision practitioners. We identify key aspects of the effective documentation of responsibilities, decision-making, and power asymmetries that decisively shape image datasets. Our investigation is framed by the following research questions: (RQ1) How can the specific contexts that inform the production of image datasets be made explicit in documentation? (RQ2) Which factors hinder documentation in this space? (RQ3) How can documentation be incentivized?

Given the complex interweaving of actors, iteration, and responsibilities involved, documenting the context of data transformations is crucial, yet hard to achieve. We propose *reflexivity*, understood as the consideration of social and intellectual factors that predetermine and shape praxis [7], as a crucial component for retrieving and documenting power dynamics in data creation. We borrow Bourdieu's "Invitation to Reflexive Sociology" [8] and translate it into an invitation to reflexive data practices. Our invitation regards reflexivity not as personal introspection but as a collective and collaborative endeavor [8].

We start by reviewing work that investigates the documentation of machine learning datasets and models. Then, we explore different conceptualizations of reflexivity. After offering an overview of research methods, informants, and fieldwork sites, we present our findings. These are organized around four salient documentation-related issues emerging from our analysis, namely *the variety of actors involved and the collaboration among them, the different purposes and forms of documentation, the perception of documentation as burden, and problems around the intelligibility of documentation.* Next, we discuss the implications of our findings and propose the implementation of reflexivity in disclosure documents for computer vision datasets. Finally, we introduce and discuss four motivations which could lead companies to implement reflexivity-driven documentation, namely, *preservation of knowledge, inter-organizational accountability, auditability, and regulatory intervention.*

2 RELATED WORK

2.1 Documentation of Datasets and Models

Previous work has pointed at the need for opening black-box algorithms by explicating their outcomes [37, 44] and documenting their modeling [26, 34]. A growing body of literature has investigated and developed structured disclosure documents or checklists for artificial intelligence models and services, which document their intended uses, testing methodologies and outcomes, actors involved, possible bias, and ethical problems [3, 15, 26, 34]. While these disclosure documents primarily focus on AI models and services, information relevant to training datasets is also required to be reported.

Recent research [4, 22, 23, 27] has called for applying similar structured procedures for documenting datasets specifically. This line of research advocates for and applies the systematic documentation of datasets' purpose, composition, collection process, preprocessing, uses, distribution [22, 23, 47], and maintenance [11, 22, 27, 47]. Several studies also draw special attention to the documentation of actors involved, including their characteristics and roles [4, 22, 23], the use of software and other tools [4, 22, 47], availability of training and additional resources for documentation [4, 23], and fair pay for workers [22, 23, 47]. Furthermore, ethical concerns have been raised in documentation regarding privacy [22, 27, 47] and potential harms of datasets [22, 47] (see Table 1).

Most prominently, Gebru et al [22] argue that documentation can improve transparency, accountability and reproducibility, and facilitate the communication between dataset consumers and producers. They propose that every dataset be accompanied by a checklist which should be flexible enough to accommodate specific domains and "existing organizational infrastructure and workflows" [22]. Holland et al. [27] argue that documentation of datasets can enable consumers to select appropriate datasets better and, at the same time, improve data collection practices among dataset creators, as they would need to explain and justify their practices. They propose a dataset nutrition label that is composed of modules to be filled in through a combination of manual work and automated procedures. Geiger et al. [23] focus primarily on documentation of datasets in academic settings. They maintain that documentation not only contributes to increasing reproducibility and open science, but is also a matter of "research validity and integrity" [23]. The authors highlight the role of major institutions to promote dataset documentation and note the technical tools and resources that can facilitate processes. Bender and Friedman [4] further suggest that such documentation should be required from "academics, industry, and government."

Whereas current proposals and practices of documentation often prioritize reproducibility, power imbalances in contexts of data creation are not often accounted for. In their investigation of data annotation services, Miceli et al. [32] present evidence of how power asymmetries shape computer vision datasets. In particular, the authors show how the judgements of managers and, even more, of requesters remain unquestioned when it comes to interpreting and labeling data. In view of these dynamics, D'Ignazio and Klein [19] underline the importance of restoring the context where datasets are produced, be it "social, cultural, historical, institutional, (...) [or] material," and the identities of dataset creators. They explain that "one feminist strategy for considering context is to consider the cooking process that produces 'raw' "data" [19] and propose asking "who questions" to drive reflection and analysis on power and privilege. In line with this research, we highlight the importance of looking into processes of data creation and foster disclosure documents that go beyond datasets' technical features. We argue that the dimensions proposed or applied in structured dataset documentation formats (see Table 1) are necessary but insufficient to drive a much-needed reflection of industry practitioners' and researchers' position and influence on data. For such a reflection to be possible, datasets must be placed in the context of their production. This perspective would not only provide a better understanding of datasets' "functional limitations" but can also make power asymmetries in data settings [19] visible.

2.2 The Notion of Reflexivity

According to D'Ignazio & Klein [19], reflexivity is a precondition for restoring context in data creation. The authors define reflexivity as "the ability to reflect on and take responsibility for one's own position within the multiple, intersecting dimensions of the matrix of domination" [19]. The matrix of domination is a concept first termed by Patricia Hill Collins [13] to explain how systems of power are configured and experienced. Black feminist scholars and critical race theorists have given considerable attention to the importance of one's positionality with regard to race, gender, and class in scientific practice. The work of Dorothy Smith [48], Patricia Hill Collins [13], and Sandra Harding [25] in standpoint theory is an important strand in this space. Researchers in critical race theory further interrogate ideological positioning of privileged and dominant groups [2, 6, 18]. More broadly, scholars on positionality frame actors' positions in socio-political contexts and scrutinize researchers' personal identities and stances concerning the contexts of knowledge and study [9, 12, 31]. These positions shape researchers' view of the world and thereby the whole research process, i.e., how they perceive, construct and approach a research problem, how they report research findings, and the process of knowledge construction and production [9, 12].

Previous investigations in sociotechnical systems have introduced reflexivity by drawing experiences and methodologies from other disciplines to examine presumptions and taken-for-granted practices in machine learning and data science. Viewing machine learning via computational ethnography, Elish and boyd [20] underline the situated nature of knowledge work and argue in favor of methodological reflections and reflexive practices. Drawing on critical race methodologies and operationalization of race in other disciplines, Hanna and Denton et al. [24] argue that the widespread conception and operationalization of race in algorithmic systems as a fixed attribute is decontextualized and, therefore, problematic. Previous work has furthermore argued that machine learning systems have positionality. Among other factors, "they inherit positionality from data" [1]. Preconceptions and values get embedded in data, for instance, through collection and analysis methods and through the taxonomies used in data annotation. The sensemaking and classification of data through labels as performed by annotators [32] is "a judgement and as such informed by the knowledge, experiences, perspectives, and value commitments of annotators or labelers" [1].

As we will explain in Discussion, Pierre Bourdieu's conceptualization of reflexivity, understood as a relational construct and an integral part of inquiry praxis, is at the core of the documentation framework we present in this paper. Bourdieu's writings on reflexivity offer a systematic investigation into social and intellectual factors that predetermine and shape researchers' practices in scientific work [7, 8, 21]. The Bourdieusian notion of reflexivity goes beyond personal experiences and regards researchers' position at the collective level, that is, in relation to other actors and the field of inquiry as a whole. Moreover, Bourdieu's reflexivity does not aim to undermine objectivity. Instead, it is presented as an analytical tool to sensitize researchers to "the social and intellectual unconscious" that condition their thoughts and practices in research, and is, therefore, an integral part of and a "necessary prerequisite" for scientific inquiry. The French sociologist pinpoints three types of bias that may influence scientific research, which may be mitigated by introducing reflexivity.*The first bias* results from researchers' positions in the social structure, such as class, gender, and ethnicity. *The second bias* comes from researchers' position in academic disciplines, i.e., academic traditions, prevailing currents, and socio-organizational structures in specific disciplines that determine specific field epistemologies. *The third bias*, termed by Bourdieu as the intellectualist bias, is embedded in the scholarly gaze that places researchers outside or above the object of research and considers their engagement with problems as purely scientific and unconstrained from social positions and economic interests. In opposition to this idea, Bourdieu argues that researchers are participants rather than external observers and restores research practices as knowledge-producing activities rather than pure and disinterested investigations. In the Discussion section, we will come back to this notion of reflexivity. The three Bourdieusian levels of bias will be the base to discuss why reflexivity is fundamental for documenting data practices. Reflexivity to make individual and collective positions explicit and acknowledge their effects on data is not only crucial for conducting better science, as Bourdieu [8] argues. It could also help researchers and practitioners uncover broader ethical issues in computer vision systems.

3 METHOD

3.1 Data Collection

This investigation was organized around two phases, involving different (yet related) research foci and methods. Documentation practices are a critical aspect we investigated at both stages:

In the first phase, we focused on work practices in data processing companies, where human workers collect, segment, and label image training data. We conducted ethnographic fieldwork at two data processing companies of the "impact sourcing" sector located in Buenos Aires, Argentina, and Sofia, Bulgaria. Impact sourcing refers to a special type of business outsourcing processing company that intentionally employs workers from marginalized communities. As described on their websites and confirmed by our observations, the Argentine company employs young people living in slums, while the Bulgarian organization works with refugees from the Middle East.

The Buenos Aires-located company that we will call "Emérita" is a medium-sized organization. With branches in three Latin American countries, Emérita conducts projects in data annotation, content moderation, and software testing. Its clients are large regional corporations in diverse fields such as security, e-commerce, and energy. At the time of the observations, between May and June 2019, the

Table 1: Summary of descriptive dimensions in documentation frameworks proposed or applied in previous research. It should be noted that the dimensions are often interconnected and not mutually exclusive.

Descriptive Dimensions in Documentation	Authors / Proposed or Applied Documentation Form					
	Gebru et al. [22]: Datasheets	Geiger et al. [23]: manual and technology-assisted documentation	Bender and Friedman[4]: Data statements	Holland et al. [27]: Dataset Nutrition Label	Seck et al. [47]: Datasheets	Choi et al. [11]: Datasheets
Description of dataset's motivation: private or public? single use or open dataset?	✓		✓		✓	✓
Description of actors involved: e.g. funding providers, data workers, data subjects and so on	✓	✓	✓		✓	✓
Description of dataset's composition	✓		✓	✓	✓	✓
Description of dataset's collection process	✓		✓	✓	✓	✓
Account of data (pre-)processing steps (e.g., cleaning,labeling)	✓	✓	✓		✓	✓
Description of dataset's intended and recommended uses	✓			✓	✓	✓
Description of datasets' distribution	✓	✓		✓	✓	✓
Description of datasets' maintenance	✓			✓	✓	✓
Description of software and other tools used in data work	✓		✓		✓	
Reflection on potential impacts and ethical issues relevant to datasets	✓			✓	✓	✓
Description of training for data workers		✓	✓			
Formal definitions and instructions for annotation		✓	✓			
Payment for workers	✓	✓			✓	
Team composition and diversity	✓	✓	✓		✓	
Account for production settings and hierarchies		✓				
Procedures for solving discrepencies in data production		✓				
Rationale for data collection framing and labeling taxonomies			✓			

Buenos Aires branch of Emérita had around 200 data-related employees who mostly worked 4 hours shifts, Mondays to Fridays, and were paid at the minimum wage.

"Action Data" is the code-name of the Bulgarian company. Action Data specializes in image data collection, segmentation, and labeling. Its clients are computer vision companies, mostly located in North America and western Europe. The company offers its workers contractor-based work and the possibility to complete their assignments remotely, with flexible hours. Contractors are paid per picture or annotation, and payment varies according to each project and its difficulty. At the time of the observations, in July 2019, the Bulgarian company was very small in size. Three employees in salaried positions and a pool of around 60 contractors handled operations.

At both sites, we conducted several weeks of observations, with different levels of interaction and involvement. All tasks observed were related to the production of datasets for computer vision and requested by computer vision companies. Moreover, we observed the on-boarding, briefing, and further training of workers as well as instances of communication between managers and teams, and managers and requesters. It is important to mention that the observations were primary conducted with a different research question in mind and focused on general work practices and not specifically on documentation. However, the exploratory character of the method and the rich interactions observed allowed us to extract useful insights for this investigation that were later corroborated by our interview partners.

In addition to the observations, fieldwork at both sites also consisted of intensively interviewing data collectors, annotators, and management. In total, we conducted sixteen in-depth interviews with an average length of 65 minutes, face-to-face, at both locations. Informants were aged 21 to 40. Eleven of them identified as female and four as male. None of them had received an education in tech-related fields or had technical knowledge prior to their current employment. At Emérita in Argentina, we conducted five in-depth interviews with data workers and employees in managerial positions. At Action Data, we conducted eleven in-depth interviews with workers and managers. Interview partners were asked to choose code names to preserve their identity and that of related informants. The interviews included accounts of specific work situations involving the interpretation of data, the communication with managers and clients, and the documentation of responsibilities and decisions. Moreover, the interviews covered task descriptions, general views on the company and the work, informant's professional and educational background, expectations for the future, and biographical details.

The second phase of this investigation dealt with the role of stakeholders at the opposite end of the service relationship, namely, the computer vision companies requesting data processing services. At fieldwork, we observed that requesters have a major influence on the documentation practice of data processing companies and decided to pursue this line of inquiry. Through expert interviews with computer vision engineers, data quality analysts, and managers, we investigated how task instructions are formulated and communicated to data processing workers, and how this process is documented. The interviews revolved around the object, purpose, and responsibilities of documentation. Moreover, we discussed issues and possible solutions for implementing broader forms of documentation in industrial contexts at the intersection of data processing and computer vision.

We conducted a total of fourteen expert interviews. Four informants were managers with large data processing companies located in Kenya, India, and Iraq. In addition, six expert interviews were conducted with computer vision practitioners working on products including an aesthetics model that sorts and rates personal image libraries, a scanner that detects contamination on hands, and optical sorting equipment for the classification of waste. The computer vision practitioners work for companies located in Germany, Spain, and the United States. Finally, four of the interviews conducted at Emérita and Action Data revolved almost exclusively around the role of requesters in documentation and were framed as expert interviews.

While the goal of in-depth interviews is revealing practices and perceptions, the purpose of expert interviews is to obtain additional professional assessments on the research topic [29]. The sampled interview partners were considered experts because they were able to provide unique insights into widespread routines and practices in their and other companies. With an average length of 48 minutes and conducted face-to-face or remotely, the expert interviews allowed us to contextualize some of the practices observed at fieldwork and analize to what extent observations could be generalized to other settings.

3.2 Data Analysis

For the analysis, we integrated field notes with a total of thirty interview transcriptions and used constructivist grounded theory principles [10] to code and interpret the data. We conducted phases of open, axial, and selective coding and let the categories emerge from the data. We applied a set of premises [14] to make links between categories visible and make them explicit in our research documentation and in open discussions among three coders. We constantly compared the collected data to revise our emergent understanding or find additional evidence of observed phenomena. Four salient axial dimensions identified during the analysis process constitute the base for the findings we present in the following section.

4 FINDINGS

As stated in Introduction, this paper explores three research questions: (RQ1) How can the specific contexts that inform the production of image datasets be made explicit in documentation? (RQ2) Which factors hinder documentation in this space? (RQ3) How can documentation be incentivized? Our findings unpack documentation practices at the intersection of data collection, data annotation, and computer vision engineering. Through descriptions and interview excerpts, we describe salient dimensions emerging from our data: *actors and collaboration, documentation purpose, documentation as burden,* and *intelligibility of documentation.* These four dimensions reveal scenarios that should be taken into account for creating effective documentation procedures that are based on workers' needs and possibilities.

4.1 Actors and Collaboration

Our first research question inquires about ways of making the specific production contexts of image datasets explicit in documentation. In this section, we take a first step towards unpacking RQ1 by describing the characteristics of such production contexts.

The creation of computer vision datasets requires the collaboration of actors that often work in different organizations. At the intersection of data collection, data annotation, and computer vision engineering, not every actor has the same influence on data [32]. Power differentials become evident when deciding which data to collect, how to classify it, and how to label it. Many datasets are produced with a specific computer vision product in mind. Dataset design begins as the expected outcome of that product (in terms of computational output but also of revenue) is transformed into task instructions for data collectors and annotators. A typical assignments is illustrated by a data collection project of Active Data: the company received task instructions to collect images of diverse human faces from a Western European company, producing identification and verification systems. Eva, the founder of Active Data, offered more details:

> "They were interested in a diversity of five different ethnicities, so Caucasian, African, Middle Eastern, Latin American and Asian. Of course, very debatable whether these can be the five categories that can classify people around the world "

This type of assignment generally revolves around a client's envisioned computer vision product and underlying business idea. The technical assumptions of a classification system demand mutually exclusive categories, in this case even for a problematic concept such as race. Whether such categorisation captures the realities of data subjects or coincides with the values and believes of data workers is not negotiated. Written instructions formulated by the requester are passed along to project managers who brief workers. Workers then start collecting the images. For outsourcing companies, the rationale behind data-related decisions is "doing what the client ordered" and "offering value to the client." Conversely, the rationale shaping datasets in computer vision companies is "data needs to fit the model" and "data processing should be fast, cost-efficient, and high-quality."

Power differentials between service providers and requesters become even more evident given that the data processing companies participating in this investigation are located in developing countries, while their clients are in the Global North. In view of such asymmetries, decisions about what to document and the financial means to do so largely depend on the most powerful actors. Anna, an intern working at Action Data and in charge of auditing the company and conducting an impact assessment, concisely described these dynamics:

> Q: "What do you think are the potential drivers or reasons for the implementation of the more transparent approach to documenting systems and processes?"
> A: "If the customer demands it."
> Q: "Is this something you have heard before, customers demanding a more ..."
> A: "No."

Moreover, computer vision companies often regard some of the information that could or should be documented as confidential, especially if it involves details about the intended product or if some of the processes involved in producing the dataset are considered a strategic advantage. Given the collaborative nature of data creation, one stakeholder's opacity may affect others' inclination towards transparency. As Active Data's founder Eva (and several others of our informants) described, secrecy in computer vision hinders her company's attempts to document work processes:

> "It's also a small challenge of how to preserve some of the know-how throughout the different projects without of course revealing too much about the different processes that each client has, you know, the confidential information from each project."

In many cases, this issue leads to reluctance to share existing documentation with other stakeholders and the general public or to not document at all.

4.2 Documentation Purpose

The reasons for documenting the production of datasets and the forms of documentation vary with each organization. To start considering ways of incentivizing documentation (RQ3), we first must look into common needs and goals that different stakeholders may have in relation to disclosure documents. In this sense, we have identified four common documentation purposes: *preservation of knowledge, improvement of work practices, accountability,* and *disclosure of dataset's specifications.*

All data processing companies participating in this investigation carry out some form of project documentation. In a more or less structured way, companies document task instructions provided by clients. Instructions may change as projects develop, or workers might develop new practices according to clients' feedback. Soo is a project manager at the Kenyan branch of a large data processing company. During our interview, he explained how this form of documentation can help improve existing processes and practices:

> "We have a 'lessons learned'- folder where we put all these items. Like the client has said, 'You did not do well here.' We'll find in our process, there was this flaw. We will document that. And then what happens after we document is that information is stored to be used for that project and some future projects with the same kind of process work."

The *preservation of this form of praxis-based knowledge* is crucial because it helps organizations resolve doubts that might emerge, train future workers, and apply situated solutions to future projects. Similarly, documentation can also serve to *revise and improve work practices and flows*, as further described by Soo:

> "How can we improve this process? This did not go well. What was the issue? How did we solve it? How can we avoid this in future? And you will get information for a project that was done five years ago [...] The documentation helps us in making sure that we avoid repeating the same mistakes. And also, it helps us in looking for better ways of doing the work, how to measure where it is possible and also what other process we can improve, like in the process flow"

Given the differentials of power described in the previous section, documentation is many times perceived as useful for *accountability* between outsourcers and requeters. Several informants working at data processing companies highlight the importance of preserving task instructions and documenting changes instructed by clients. Keeping this type of record might serve as proof that tasks were carried out as instructed. In the next interview excerpt, the founder of Active Data describes how documentation might help resolve discrepancies if clients are not satisfied with the quality of the service provided or decide to demand more:

> "We also keep the client accountable so that they don't come up with a new requirement or something that we haven't mentioned before. So, SoWs [scope of work documents] are also for accountability of us towards the client as well so that the client can have a document where they can keep track of what the arrangement is and so on beyond our contract"

However, accountability within teams can become surveillance for workers: several informants account for the connection between project documentation and the measurement of workers' performance in data processing companies. The Argentine company, Emérita, directs great efforts to measure workers' performance and output quality and to transform those into numbers and charts. Nati, Emérita's continuous improvement analyst, described this process:

> "Within the project documentation, we have an external person who checks if the work the team did is right or wrong, then documents the percentage of right and wrong. [...] If something is wrong, we fix it before the client notices. But still, even when it is fixed, we record that there was something that was wrong and record who was responsible for the mistake."

Finally, in the case of datasets for public use or without a pre-established purpose, organizations might find it important to document and disclose *datasets' specifications.* This particular case was reported by our informants at Action Data, as the company had recently released two datasets for public use. During an interview, Eva contemplated the possibility of releasing a disclosure document along with the datasets:

> "It might be nice to implement some type of documentation at least for them [datasets for open use] because they're for external use and it might be good to know what the origin of the images are, what the process of annotation had been and so on."

It is worth mentioning that releasing datasets for public use is usually not within the scope of outsourcing companies. Investing resources to produce a pro-bono dataset represents a considerable effort for these companies. In the case of Active Data, the dataset was made publicly available as part of the company's marketing strategy.

4.3 Documentation as Burden

Relevant to start unpacking factors that hinder documentation (RQ2) is the fact that several informants see documentation as time-consuming, extra work that is likely to delay the completion of workers' "actual" tasks. This is a widespread view among the computer vision practitioners interviewed for this investigation and coincides with the observation that, among the different roles explored in this study, computer vision companies seem to be the least inclined to document work practices.

"Lack of time" is the most widespread answer when informants are asked why there are not more aspects of data creation reflected in reports. Documentation is broadly perceived as optional, a nice-to-have feature that is implemented only once all "important" issues are sorted. Andre, a US-based computer vision engineer with a start-up dedicated to producing scanners that detect contamination on hands, described his company's position on this issue:

> "[Documenting] is lower on our priority list than a bunch of other things that we need to do. It's just not the company's priority at this moment. There are other more valuable things to keep the company successful. As the engineering team grows, as we have more time to do those things and our work to meet the company's exact needs are less burdensome, then we'd go to more documentation."

Among our informants in computer vision companies, the view persists that documentation is an activity only large corporations can afford. As further reported by Andre, start-up teams are smaller, and workers are multitasking, which reinforces the view that there are more pressing issues than documentation:

> "That's one of the interesting things about start-ups. You don't have the time to document everything. [...] There is a lot of knowledge in every single person here that would take far too long to pull out of them and transfer to a new person and keep the company still running at the same time."

A similar observation was made by Eva, the founder of the Bulgarian data processing company, regarding her company's clients:

> "We've been working with quite a lot of new companies recently. Some of them are bigger corporations that have more let's say bureaucratic procedures and more detailed processes of description of everything that's happening around the project, while others are just start-ups that prefer very lightweight, minimum involvement and paperwork around their projects."

Lack of incentives, external or internal, is another reason why documentation might be perceived as a burden. For instance, some informants agreed that laws and regulations would be an excellent external incentive for technology companies to integrate documentation as a constitutive part of their work. In the absence of regulations, documentation is seen as optional extra work. As for internal incentives within organizations, several computer vision practitioners explained that documenting was not a part of their work routines and was therefore not encouraged by the company's structures. Emmanuel, a computer vision engineer based in Barcelona and working on optical sorting equipment for waste's classification, discussed the need for integrating documentation in existing workflows. He moreover imagines that extending projects' deadlines to prioritize documentation would not be seen as acceptable within his company's culture:

> "Time is a huge issue. I mean, I think planning is very important, get the time to do it [documenting] and that everybody knows this is supposed to be done. Because right now, documenting is not a task and I don't know that I would have a gap between projects so I could document. And this is never a priority for the company, they expect me to meet my deadlines, I can't just drop my deadlines to document. And this is a problem. If documenting was part of the deadline, companies wouldn't just leave it for another time"

Even in companies that integrate laborious documentation in their work processes, as is the case of Emérita, there are instances where documenting is just not profitable. Nati, one of our informants with the Argentine data processing company, describes one of those situations:

> "It happens sometimes that we do one-time projects that go only for one or two weeks. In those cases, documentation is a waste of time and money, because the client buys, let's say, eighty hours and you spend twenty documenting. It's just not profitable."

As expected, financial incentives, or the lack thereof, can also influence views on documentation.

4.4 Intelligibility of Documentation

To further investigate factors that hinder documentation (RQ2) it is necessary to explore issues around creating compelling, retrievable, and intelligible disclosure documents. To illustrate some relevant aspects related to structuring and providing access to documentation, we draw on the observations made during fieldwork at both data processing companies, Emérita and Active Data. Both companies have vast experience in the documentation of data collection and annotation projects.

In the case of the Argentine company, Emérita, due to the extension of documentation and the large number of projects conducted, navigating and maintaining disclosure documents has become difficult. Nati, a continuous improvement analyst, is in charge of addressing this issue:

> "What happened a lot was that information was repeated in many places. The objectives were written in three different documents. The people who were in the project were in two different systems [...] So, having that repeated was horrible, because every time people in the team changed, well, you needed to update many things and credentials"

Nati works on optimizing some of her company's internal processes, including documentation. For that purpose, she has surveyed project documents, observed how the company teams work, and discussed with them how documentation can be improved. Her main focus lies in producing documentation that can be easily retrieved and used, which can be very challenging:

> "For example, in the case of project guides, it was not clear what documentation had to be done, so everyone did what they wanted, or what they remembered, or what they knew, because someone told them, and when information was needed, they didn't know if it had been documented or not, or they didn't know where to find it. We lost a lot of information like this."

Further issues related to the intelligibility of documentation may arise depending on who is in charge of documenting and who are the users of documentation. In the case of Active Data, the Bulgarian company working with refugees from the Middle East, language and lack of technical knowledge is one of those issues:

> "Since we're working with people who very frequently do not have high levels of education or do not speak good English, I've heard a lot of complaints that people are not reading the training documents or they're not following them or they're asking questions that appear or are already answered in the training documents. So, it can be quite frustrating because people may not be used to following such documentation and they might need additional training just to know how to use this recommendation, how to read it and how to follow it"

Creating useful reports that can be easily retrieved and understood is challenging. How disclosure documents are created, indexed, and stored depends to a greater extent on the intended addressees of documentation. As illustrated by the previous interview excerpt, language is important if stakeholders with different levels of literacy will make use of documentation.

5 DISCUSSION

As described in Findings, work at the intersection of data collection, annotation, and computer vision engineering requires strong coordination efforts among actors that occupy different (social) positions. Documentation purpose, organizational priorities, and needs around documentation intelligibility vary across stakeholders. In such heterogeneous contexts, some actors hold more power than others and decisions made at the most powerful end will inevitably affect work practices and outputs at every level. These power differentials and their effects are broadly naturalized [17, 19, 32]. Despite their decisive effects on data, decisions and instructions that are rooted in such naturalized power imbalances are mostly perceived as self-evident and remain undocumented as a consequence.

Previous research has emphasized the importance of documenting machine learning datasets [22, 23, 27, 30, 49]. While we acknowledge that work for creating the foundations for our investigation, we also argue that the frameworks proposed are not sufficient to interrogate power differentials and naturalized preconceptions encoded in data. With our investigation, we move the focus away from documenting datasets' technical features and highlight the importance of accounting for production contexts. Our research questions address the challenge of documenting production processes that are characterized by the multiplicity of actors, needs, and decision-making power. In this and the following sections, we lay out implications of our observations and outline a documentation framework to address the contexts and issues described in Findings.

Given the collaborative nature of datasets production, we argue that documentation should not be carried out in the vacuum of each organization. The framework we propose regards dataset documentation as a collaborative project involving all actors participating in the production chain. This is not easy for sure. To address such challenge, we propose that reflexivity, understood as a collective endeavor [7], be an integral part of such collaborative documentation. As argued by Bourdieu [8], this form of collective reflexivity accounts for actors' social position and aims to interrogate praxis fields and the relations that constitute them. In a similar manner, reflexive documentation should help to make visible the interpersonal and inter-organizational relations that shape datasets. As described in the Related Work section, Bourdieu's notion of reflexivity covers three levels of hidden presupposition: the researcher's social position, the epistemology of each disciplinary field, and "the intellectualist bias", described as the scholarly gaze researchers use to analyze the social world as if they were not part of it [7, 8]. We take this perspective and transform Bourdieu's "Invitation to Reflexive Sociology" [8] into an invitation to reflexive data practices. What constitutes our invitation entails much more than observing how one actors' positionality affects data: If documentation is to be seen as a collaborative project, reflexivity of work practices should be understood as a collective endeavor, where widespread assumptions, field methodologies, and power relations are interrogated.

With this framework, we regard documentation in a two-fold manner: First, as an artifact (the resulting documentation) that enables permanent exchange among stakeholders participating in data creation. We envision disclosure documents that travel among actors and organizations, across cultural, social, and professional boundaries, and are able to ease communication and promote inter-organizational accountability. Second, we regard documentation as a set of reflexive practices (the act of documenting) intended to make naturalized preconceptions and routines explicit. Just as Bourdieu regards reflexivity as a "necessary prerequisite" for scientific inquiry, the reflexive practices involved in our documentation framework should be seen as a constitutive part of data work. If reflexivity is only regarded as a desirable goal related to AI ethics and not as actual *part of the job*, documentation will never be considered a priority and, as described in Findings, it will continue to be perceived as a burden.

5.1 Why Reflexivity?

Our research questions enquire about ways of making the contexts that inform the production of image datasets explict in documentation and about factors that hinder or incentivize the implementation of documentation in industry settings. In view of our findings, we argue that effective documentation should be able to reflect the dynamics of power and negotiation shaping datasets through work practices. However, making visible the hierarchies, worldviews, and interests driving decisions and instructions is extremely challenging. One major difficulty lies in their taken-for-grantedness: documenting naturalized power dynamics and decisions that are largely perceived as self-evident [33] require intensive reflexive practice.

The three previously-mentioned levels of reflexivity proposed by Bourdieu (social position, field epistemology, and intellectualist gaze) can be useful to discuss why reflexivity should be at the core of documentation practices in data creation for computer vision. They provide an additional lens through which data practices can be approached, and as such, serve as a complement to on-going work and discussions regarding the documentation of datasets:

First, reflexive documentation should consider the social position of workers involved in dataset production, not just individually but in their relation to other stakeholders. Such consideration could help produce documentation that brings power imbalances into light and questions taken-for-granted instructions and hierarchies. This relational examination is especially important due to the widespread use of outsourced services for the collection and annotation of data: Workers at crowdsourcing platforms are subject to precarious employment conditions [28, 45]. In the impact sourcing companies presented in this paper, workers come from marginalized communities (refugees in Active Data, slum residents in Emérita). Most of them have no technical education. How does their social position affect these workers' ability and power to question the instructions commanded by computer vision engineers or data scientists in tech companies? This question becomes even more pressing if we examine the relationship that connects data processing services in developing countries with computer vision companies in the Global North. Documentation frameworks that are oblivious to the fact that production chains are shaped by asymmetrical relationships will never be effective in reflecting how those asymmetries affect data. In this sense, reflexive documentation should bring power differentials to light and, ideally, empower those in vulnerable positions to speak up and raise questions.

Second, reflexive documentation should serve to question field epistemologies. Examining the epistemology of computer vision might shed light on the assumptions, methods, and framings underlying the production of image datasets. As Crawford and Paglen [16] argue, computer vision is "built on a foundation of unsubstantiated and unstable epistemological and metaphysical assumptions about the nature of images, labels, categorization, and representation." Bringing these assumptions forward in documentation is important because socially-constructed categories, such as race and gender, are generally presented as indisputable in image datasets [46]. Furthermore, a fixed and universal nature is not only ascribed to the categories as such, but also to the correspondence that supposedly exists between images and categories, appearances and essences [16]. Reflexivity should help reveal the political work such assumptions perform behind their purely technical appearance.

Finally, reflexive documentation should help practitioners question the "intellectualist gaze" [7] in data work. This type of bias is the inclination to place ourselves outside the object of research. This form of examination would highlight the role of workers and organizations in creating data while questioning widespread notions such as "raw data" and "ground truth labels". Reflexivity should therefore help to adopt a relational view on data and data work, acknowledging data as a "human-influenced entity" [35] that is shaped by individual discretion, (inter-)organizational routines, and power dynamics.

5.2 Why Document?

Data processing services and computer vision companies might be reluctant to implement such an elaborate approach to documentation. Our third research question asks how can documentation be incentivized. In this section, we consider four ways in which the Bourdieusian framework previously outlined can constitute an asset for organizations, and thus serve as an incentive for the uptake of reflexive documentation.

5.2.1 *Preservation of Knowledge.* Reflexive documentation could make praxis-based and situated decision-making explicit and help preserve it in documentation. This knowledge can become long term business assets for companies. Moreover, reflexive documentation can preserve know-how relevant to data work [39] that may get lost due to workers flow. As the flow of workers brings about problems in task transfer and reinvestment in training new employees, documentation that preserves knowledge and methods for effective data work, be they project-specific or not, can ease the transition.

Furthermore, documentation can "have analytical value [and] improve communication in interdisciplinary teams" [32]. The framework offered in this paper highlights the collective nature of reflexivity. We argue that documentation that preserves praxis-based knowledge and best practices (as described in section 4.2) should be circulated among collaborating companies rather than be produced and retrieved in the vacuum of each organization. For one thing, sharing such documentation with other stakeholders may improve the quality of data work and of the datasets that are produced as a result. For another, documentation providing more details on discretionary decision-making and its contexts can enhance transparency and facilitate a better understanding of datasets before model development.

5.2.2 *Inter-organizational accountability.* Tracking decisions and responsibilities in environments and processes that involve multiple organizations can be challenging. As described in Findings, data processing companies use documentation to foster inter-organizational accountability and protect themselves in the face of disagreements with clients. At the same time, computer vision companies might consider documentation as a tool to keep track of the processing status of projects and audit requested tasks. Reflexive documentation could be especially useful to improve traceability, as the participation of many actors and iterations in data creation may lead to accountability dilution [32]. Moreover, documentation could provide "organizational infrastructure" that empowers individual advocates among workers to raise concerns and reduces the social costs for such actions [30]. An infrastructure based on the reflexivity framework outlined in this paper could facilitate the interrogation of intra- and inter-organizational relations, normative assumptions, and workflows shaping data at the three levels described in the previous section.

Conducting documentation at a collaborative level, which means to engage various actors and to accommodate documentation to their needs, can serve as a platform for permanent exchange among stakeholders. Enabling permanent exchange could help anticipate disagreements and misunderstandings, thus improving task quality and reducing completion time.

5.2.3 Auditability. Documentation based on reflexivity could constitute an asset for organizations to prevent issues before they are made public or weather the storm in the face of PR failures. Disclosure documents that are able to retrieve the context of dataset production could constitute a useful tool for auditability, for instance, when computer vision outputs are publicly questioned or for internal ethics teams who would like to perform an assessment for potential fairness concerns prior to the release of a model trained on such data [42, 43]. Such documents could help to identify problematic issues before they become public pushbacks. Moreover, in case of public failures, documentation could provide an audit trail that would allow organizations to address problems and offer solutions promptly. In this sense, public pressure could constitute an incentive for companies towards documentation.

In such cases, counting with reflexive documentation to audit datasets could help companies offer solutions that go beyond "throwing in more data" and are able to address issues at the three Bourdieusian levels previously described: identifying asymmetrical relationships that might have been encoded in datasets, interrogating widespread assumptions in computer vision, and questioning data, even "raw" data.

5.2.4 Regulatory Intervention. Organizations could also be pushed towards documentation through regulatory intervention. Yet, before any form of reflection, including the documentation thereof, can be imposed, a few observations can be made:

First, while documentation might be considered an important component or step of the reflexive process, it is neither constitutive to, nor sufficient for, reflection. Reflexivity represents a state of awareness, an encouragement for actors involved in data creation to more widely consider the impact of their practices. Reflexivity can already be valuable in itself. The policy end-goal is therefore to stimulate a reflexive mindset and to establish the right conditions for such a mindset to fully come to fruition. Conversely, if regulation only aims at pushing documentation, the danger exists that such regulatory requirements are approached as merely an administrative exercise towards compliance.

Second, if the encouragement of reflexivity through legal means would be desired, such mechanisms may already be (partially) present in existing initiatives. For instance, it could be argued that the EU General Data Protection Regulation's increased emphasis on accountability and risk-based responsibility stimulates some level of reflection where personal data are involved [36]. Reflexivity could moreover become an additional supportive tool for data workers as a means to detect and mitigate the impact data actions have on (fundamental) rights, and as such, contribute towards the compliance with existing legal frameworks.

Third, given the multiplicity of actors involved in data creation, regulatory initiatives should also carefully consider the actors they wish to target. Stakeholders should not only be targeted in isolation; instead, policy makers should understand the relationships these actors hold vis-a-vis one another, and the consequences that their relationships bear on the activities performed.

Finally, any regulatory response must adequately consider the power asymmetries described in this paper, including their manifestation within a globalized, international environment. Mechanisms of provenance, such as documentation, could help ensure and demonstrate that societal values and fundamental rights, as well as an appropriate level of reflexivity, have been maintained throughout the computer vision value chain, rather than purposefully avoided via outsourcing strategies and/or the exercise of power. Similarly, provenance may increase the accountability and responsibility of powerful entities in both their actions and the instructions they give.

6 LIMITATIONS AND FUTURE WORK

This investigation was designed to be qualitative and exploratory. Our findings are bound to the specific contexts of the companies and individuals participating in our studies and cannot be generalized to all computer vision production settings. In the future, we seek to broaden this research by investigating ways of integrating the framework outlined in this paper in real-world production workflows and co-designing actionable guidelines for reflexive documentation together with industry practitioners.

7 CONCLUSION

Based on fieldwork at two data processing companies and interviews with data collectors, annotators, managers, quality assurance analysts, and computer vision practitioners, we described widespread documentation practices and presented observations related to the purpose, challenges, and intelligibility of documentation.

In view of these findings, we proposed a reflexivity-based approach for the documentation of datasets, with a special focus on the context of their production. We described documentation as a set of reflexive practices and an artifact that enables permanent exchange among actors and organizations. We argued that disclosure documents should travel across organizational boundaries, and be able to ease communication and foster inter-organizational accountability. We imagined documentation as a collaborative project and argued that reflexivity of work practices should therefore be understood as a collective endeavor, where not only personal positions but also praxis fields are interrogated.

Achieving a healthy balance between these elements and incentivizing practitioners and organizations to implement reflexive documentation is not easy. The challenge is nevertheless worth exploring if we aim at addressing some of the ethical issues related to the production of data for computer vision systems.

8 ACKNOWLEDGMENTS

Funded by the German Federal Ministry of Education and Research (BMBF) – Nr. 16DII113f. For his contribution, Laurens Naudts has received support from the Weizenbaum Institute Research Fellowship programme. We dearly thank the individuals and organization participating in this study. Thanks to Leon Sixt, Matt Rafalow, Julian Posada, Gemma Newlands, and our anonymous reviewers for their valuable feedback. We thank Philipp Weiß for his support whenever we struggled with Overleaf. Special thanks to Prof. Bettina Berendt for her continuous support for this project.

REFERENCES

[1] Yewande Alade, Christine Kaeser-Chen, Elizabeth Dubois, Chintan Parmar, and Friederike Schüür. 2019. Towards Better Classification. (2019), 4. https://drive.google.com/file/d/14uL1DQN8hRyDDDAm2WEleYbmxP7dqP72/view

[2] Michelle Alexander. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (revised edition ed.). New Press, New York.

[3] M. Arnold, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, K.R. Varshney, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilovic, R. Nair, K. Natesan Ramamurthy, and A. Olteanu. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6:1–6:13. https://doi.org/10.1147/JRD.2019.2942288

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

[5] Bettina Berendt. 2019. AI for the Common Good?! Pitfalls, challenges, and ethics pen-testing. *Paladyn, Journal of Behavioral Robotics* 10, 1 (Jan. 2019), 44–65. https://doi.org/10.1515/pjbr-2019-0004

[6] Eduardo Bonilla-Silva. 2006. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States.* The Rowman & Littlefield Publishing Group, Inc., Lanham. OCLC: 781274997.

[7] Pierre Bourdieu. 2000. *Pascalian meditations.* Stanford University Press, Stanford, Calif. OCLC: 833852849.

[8] Pierre Bourdieu, Loïc J. D. Wacquant, and Loïc J. D. Wacquant. 1992. *An Invitation to Reflexive Sociology.* University of Chicago Press. tex.ids: bourdieu1992b googlebooksid: rs4fEHa0ijAC.

[9] B. Bourke. 2014. Positionality: Reflecting on the Research Process. *The Qualitative Report* 19, 33 (2014), 1–9. https://nsuworks.nova.edu/tqr/vol19/iss33/3

[10] Kathy Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis.* Sage Publications, London ; Thousand Oaks, Calif.

[11] Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. QuAC : Question Answering in Context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics, Brussels, Belgium, 2174–2184. https://doi.org/10.18653/v1/D18-1241

[12] David Coghlan and Mary Brydon-Miller (Eds.). 2014. *The Sage encyclopedia of action research.* SAGE Publications, Inc, Thousand Oaks, California.

[13] Patricia Hill Collins. 1990. *Black feminist thought: knowledge, consciousness, and the politics of empowerment.* Number v. 2 in Perspectives on gender. Unwin Hyman, Boston.

[14] Juliet M. Corbin and Anselm L. Strauss. 2015. *Basics of qualitative research: techniques and procedures for developing grounded theory* (fourth edition ed.). SAGE, Los Angeles. https://us.sagepub.com/en-us/nam/basics-of-qualitative-research/book235578 tex.ids: dnsc2015.

[15] Henriette Cramer, Jean Garcia-Gathright, Sravana Reddy, Aaron Springer, and Romain Takeo Bouyer. 2019. Translation, Tracks & Data: An Algorithmic Bias Effort in Practice. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19).* ACM, New York, NY, USA, CS21:1–CS21:8. https://doi.org/10.1145/3290607.3299057 event-place: Glasgow, Scotland Uk.

[16] Kate Crawford and Trevor Paglen. 2019. Excavating AI: The Politics of Images in Machine Learning Training Sets. https://www.excavating.ai tex.ids: zotero-3263.

[17] 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:2007.07399 [cs]* (July 2020). http://arxiv.org/abs/2007.07399 arXiv: 2007.07399.

[18] Robin J. DiAngelo. 2018. *White fragility: why it's so hard for white people to talk about racism.* Beacon Press, Boston.

[19] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data feminism.* The MIT Press, Cambridge, Massachusetts. https://mitpress.mit.edu/books/data-feminism

[20] M. C. Elish and danah boyd. 2018. Situating methods in the magic of Big Data and AI. *Communication Monographs* 85, 1 (Jan. 2018), 57–80. https://doi.org/10.1080/03637751.2017.1375130

[21] Mustafa Emirbayer and Matthew Desmond. 2012. Race and reflexivity. *Ethnic and Racial Studies* 35, 4 (April 2012), 574–599. https://doi.org/10.1080/01419870.2011.606910

[22] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. Datasheets for Datasets. *arXiv:1803.09010 [cs]* (March 2020). http://arxiv.org/abs/1803.09010 arXiv: 1803.09010.

[23] R. Stuart Geiger, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. 2020. Garbage in, garbage out? do machine learning application papers in social computing report where human-labeled training data comes from?. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20).* Association for Computing Machinery, Barcelona, Spain, 325–336. https://doi.org/10.1145/3351095.3372862

[24] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a Critical Race Methodology in Algorithmic Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20).* Association for Computing Machinery, Barcelona, Spain, 501–512. https://doi.org/10.1145/3351095.3372826 tex.ids: hanna2020a.

[25] Sandra Harding. 1993. Rethinking Standpoint Epistemology: What is "Strong Objectivity"? In *Feminist Epistemologies.* Routledge, 49–82.

[26] Michael Hind, Stephanie Houde, Jacquelyn Martino, Aleksandra Mojsilovic, David Piorkowski, John Richards, and Kush R. Varshney. 2020. Experiences with Improving the Transparency of AI Models and Services. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20).* Association for Computing Machinery, 1–8. https://doi.org/10.1145/3334480.3383051

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

[28] 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 (CHI '13).* Association for Computing Machinery, Paris, France, 611–620. https://doi.org/10.1145/2470654.2470742

[29] Natalia M Libakova and Ekaterina A Sertakova. 2015. The Method of Expert Interview as an Effective Research Procedure of Studying the Indigenous Peoples of the North. *Journal of Siberian Federal University. Humanities & Social Sciences* 8, 1 (2015), 114–129. https://doi.org/10.17516/1997-1370-2015-8-1-114-129

[30] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).* Association for Computing Machinery, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376445 tex.ids: madaio2020a.

[31] Frances A. Maher and Mary Kay Tetreault. 1993. Frames of Positionality: Constructing Meaningful Dialogues about Gender and Race. *Anthropological Quarterly* 66, 3 (1993), 118–126. https://doi.org/10.2307/3317515

[32] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (2020), 25. https://doi.org/10.1145/3415186

[33] Milagros Miceli, Martin Schüßler, and Tianling Yang. 2020. Between Subjectivity and Imposition: A Grounded Theory Investigation into Data Annotation. (2020), 19.

[34] 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 (FAT* '19).* Association for Computing Machinery, 220–229. https://doi.org/10.1145/3287560.3287596

[35] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, Glasgow, Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300356

[36] Laurens Naudts. 2019. How Machine Learning Generates Unfair Inequalities and How Data Protection Instruments May Help in Mitigating Them. In *Data Protection and Privacy : The Internet of Bodies* (first ed.), Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth, and Paul De Hert (Eds.). Hart Publishing, Oxford, 71–92.

[37] High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI.* Technical Report. European Commission.

[38] Samir Passi and Solon Barocas. 2019. Problem Formulation and Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19).* Association for Computing Machinery, Atlanta, GA, USA, 39–48. https://doi.org/10.1145/3287560.3287567

[39] 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 (CSCW '17).* Association for Computing Machinery, Portland, Oregon, USA, 2436–2447. https://doi.org/10.

1145/2998181.2998331

[40] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 1–28. https://doi.org/10.1145/3274405

[41] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and S. S. Iyengar. 2018. A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Comput. Surv.* 51, 5 (Sept. 2018), 92:1–92:36. https://doi.org/10.1145/3234150

[42] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, 429–435. https://doi.org/10.1145/3306618.3314244

[43] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 33–44. https://doi.org/10.1145/3351095.3372873

[44] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco California USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[45] Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 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 - CHI '15*. ACM Press, Seoul, Republic of Korea, 1621–1630. https://doi.org/10.1145/2702123.2702508

[46] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. 2020. How We've Taught Algorithms to See Identity: Constructing Race and Gender in Image Databases for Facial Analysis. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (2020). https://doi.org/10.1145/3392866 Article 058.

[47] Ismaïla Seck, Khouloud Dahmane, Pierre Duthon, and Gaëlle Loosli. 2018. Baselines and a datasheet for the Cerema AWP dataset. In *Conférence d'Apprentissage CAp (Conférence d'Apprentissage Francophone 2018)*. Rouen, France. https://doi.org/10.13140/RG.2.2.36360.93448

[48] Dorothy E. Smith. 1990. *The conceptual practices of power: a feminist sociology of knowledge*. Northeastern University Press, Boston.

[49] Jennifer Wortman Vaughan and Hanna Wallach. 2020. A Human-Centered Agenda for Intelligible Machine Learning. In *Machines We Trust: Getting Along with Artificial Intelligence*. http://www.jennwv.com/papers/intel-chapter.pdf