

# Wisdom for the Crowd: Discursive Power in Annotation Instructions for Computer Vision

Milagros Miceli  
Weizenbaum Institute - TU Berlin  
Germany  
m.miceli@tu-berlin.de

Julian Posada  
University of Toronto  
Canada  
julian.posada@mail.utoronto.ca

## Abstract

*Developers of computer vision algorithms outsource some of the labor involved in annotating training data through business process outsourcing companies and crowdsourcing platforms. Many data annotators are situated in the Global South and are considered independent contractors. This paper focuses on the experiences of Argentinian and Venezuelan annotation workers. Through qualitative methods, we explore the discourses encoded in the task instructions that these workers follow to annotate computer vision datasets. Our preliminary findings indicate that annotation instructions reflect worldviews imposed on workers and, through their labor, on datasets. Moreover, we observe that for-profit goals drive task instructions and that managers and algorithms make sure annotations are done according to requesters' commands. This configuration presents a form of commodified labor that perpetuates power asymmetries while reinforcing social inequalities and is compelled to reproduce them into datasets and, subsequently, in computer vision systems.*

## 1. Introduction

Computer vision systems are built from training data previously collected, cleaned, and annotated by human workers. Companies and research institutions outsource many of these tasks through online labor platforms [10] and business process outsourcing (BPO) companies [7]. In these instances, outsourcing organization and their clients regard workers as independent contractors, turning them into mere factors of production, and their labor into a commodity or a product subject to market regulations [12]. They are paid a few cents of a dollar per task, mostly lack social protection traditionally tied with employment relations, and are subject to control and surveillance [11, 4, 5]. Their assignments broadly comprise the interpretation and classification of data, and the work practices involved imply sub-

jective social and technical choices that have significant yet hidden and vaguely explored ethical and political implications. Workers interpreting and classifying data do not do so in a vacuum: their labor is embedded in large industrial structures and deeply intertwined with naturalized profit-oriented interests [6]. How much room is left for horizontal discussion and agreement among data workers regarding how data is interpreted and sorted? Can the agency of workers prevail in such environments?

In this paper, we outline our on-going research that investigates the work of data annotators in Venezuela and Argentina. We discuss how the worldviews and interests of computer vision developers permeate datasets through commodified outsourced labor. Our research is based on 62 in-depth interviews with data annotators (on the outsourcing side) and computer vision practitioners (at the requesters' end), and the qualitative analysis of 220 annotation instruction documents formulated by requesters and carried out by outsourced workers. Our preliminary findings show that, instead of seeking the "wisdom of crowds," where a diverse and independent group cooperates to solve a problem, requesters of outsourced labor impose predefined forms of interpreting, classifying, and sorting data. Requesters exert their influence through narrow task instructions and specially tailored work interfaces. Managers and quality assurance analysts in BPOs, and algorithms in online platforms are in charge of overseeing the process and making sure tasks are completed according to the needs of requesters.

## 2. Study Design and Method

Through our research, we move away from approaches that aim at "identifying" and "mitigating" bias at the individual level and propose a broader view on this issue: We explore power asymmetries in data work and their effects on training datasets. Our research questions are:

RQ1: How are ground truth labels for computer vision datasets instructed to annotators?

RQ2: What discourses are implicit in annotation instruc-

tions?

RQ3: What are the managerial practices related to those discourses?

As we describe in the following sub-sections, the exploration of these research questions comprises two qualitative methods: critical discourse analysis and in-depth interviewing.

## 2.1. Discourse Analysis

Task instructions provided by requesters to labelers play a crucial role in data annotation. They prescribe specific ways of interpreting data and the phenomena synthesized in that data. While task instructions significantly constrain the room for annotators' biases, they constitute a fundamental tool to assure the imposition of computer vision companies' values and interests on datasets. Annotators embed arbitrary preconceptions in training datasets by following said instructions.

As part of this project, we are in the process of analyzing a corpus of 220 annotation instruction documents for computer vision data. We focus on instructions for annotation tasks carried out through three data annotation platforms in Venezuela and one business process outsourcing company in Argentina. Most of the documents comprise instructions related to facial recognition systems, autonomous vehicles, and scene understanding. Our analysis focuses on the taxonomies at the core of pre-defined truth values instructed by requesters. Through critical discourse analysis, we identify naturalized discourses encoded in them. The analysis of taken-for-granted assumptions embedded in discourses is fundamental because "Discourse (is) an interrelated set of texts, and the practices of their production, dissemination, and reception, that brings an object into being [...] In other words, social reality is produced and made real through discourses, and social interactions cannot be fully understood without reference to the discourses that give them meaning." [9] Critical discourse analysis explores the meanings that become dominant and those that are marginalized. Discourses encoded in annotation instructions will determine the realities that the dataset is able to reflect and the ones that will remain erased from it.

We analyze discourses encoded in annotation instructions using a three-dimensional framework [1]: The first dimension is that of the text, and corresponds with an examination of the actual content, structure, and meaning of each instruction document. The second dimension refers to the study of discursive practices. Here, the focus is set on the form of discursive interaction used to communicate meaning and beliefs. Finally, we examine the social practice dimension, i.e., the social contexts in which task instructions are composed and carried out.

Source	Type	Content	Number
A	BPO	Various	6
B	Platform	Content Moderation	32
C	Platform	Self-Driving Cars	152
D	Platform	Search Engines	20

Table 1. Sources of instructions

## 2.2. In-depth Interviews

In addition to the discourse analysis, we have so far conducted 62 in-depth interviews with computer vision practitioners and data annotators working at crowdsourcing platforms and business process outsourcing companies. The interviews' goal is revealing practices and perceptions as well as obtaining additional information about the contexts and relations that inform how annotation tasks come to be and how instructions are communicated. The interviews include accounts of specific work situations involving the interpretation of data as well as the communication between annotators, managers, and requesters. Moreover, the interviews cover task descriptions, widespread routines and work practices as well as general views on the work and the data annotation and computer vision fields.

We use a form of constructivist grounded theory [2] to code and interpret the interview transcripts. We follow phases of open, axial, and selective coding and let the categories emerge from the data. We apply a set of premises [3] to make links between categories visible and make them explicit in our research documentation and in open discussions among the authors. We constantly compare the insights emerging from the interviews with the findings of our discourse analysis to revise our understanding or add additional evidence of observed phenomena.

## 3. Preliminary Findings

In concurrence with previous literature [7], we have observed that workers interpreting, sorting, and labeling data do not do so in the vacuum of their personal judgment: their work and subjectivities are embedded in large industrial structures and subject to control. Power asymmetries in the production of datasets manifest in decisions related to what is considered data and how each data point is interpreted. We argue that possibilities for decision and interpretation in data work are constrained through top-down approaches. Artificial intelligence politics are inextricably connected to the power relations behind the collection and transformation of data and data work conditions, which allow preconceived hegemonic forms of knowledge to get encoded in computer vision systems via training datasets.

In this case, the influence of the most powerful actors, i.e. computer vision companies and research institutions outsourcing the annotation assignments, permeates datasets

through predefined forms of interpreting and sorting data. The annotation instructions reflect the particular world-views of these actors. They require workers to categorize, label, enter data, and identify people, objects, and animals precisely, according to a set of preconceived ideas of how the world works and looks. Thus, the taxonomies present in the instructions carry meanings that are self-evident to the requesters but not necessarily relevant to the annotators and their communities. For instance, most instructions to label images according to racial categories are based in US-centered conventions. For self-driving cars, annotation instructions refer to categories present in North American roads, such as signs and animals, that make little to no sense in Latin-American contexts.

Moreover, annotation instructions reflect the commercial goals behind the computer vision systems that will be created on the base of annotated data. Predefined categories and labels mainly follow profit-oriented goals that prioritize classifications that have commercial application or are easier to operationalize in computational terms. For instance, when asked what would be the procedure if they were unsure about annotation instructions, most of the workers at the Argentine BPO answered that they would immediately seek requesters' opinion because "their interpretation is usually the one that makes more sense as they know exactly what kind of system they're developing and how they plan to commercialize it."

Furthermore, the influence of powerful actors in data annotation is stabilized through narrow task instructions, specially tailored work interfaces, managers and quality assurance analysts in BPOs, and algorithms in labor platforms. Annotation instructions make reference to such control mechanisms. Instruction documents include warnings such as "low quality responses will be banned and not paid" or "accurate responses are required. Otherwise you will be banned." These messages reinforce the annotation process's hierarchical structure and compel workers to follow norms as instructed. Data workers have little room to improve the labels or voice ethical concerns. They can only provide feedback on the aspects of the annotation process, that contribute to maintaining the status quo and power relations, such as pointing out a problem with the annotation interface or any other issues that could hinder the annotation process. Workers at the Argentine company are encouraged to think in terms of "what the client might want and what would bring more value to them". Here, any concerns expressed at the workers' end are filtered throughout hierarchical managerial structures and hardly ever reach requesters. Platform workers risk being banned and even expelled from the platform if they dare contradict task instructions.

## 4. Conclusion

While the potentially harmful effects of algorithmic biases continue to be widely discussed, it is also essential to address how power imbalances and imposed classification principles in data creation contribute to the (re)production of inequalities by computer vision systems. The empowerment of workers and the decommodification of their labor away from market dependency[10], as well as the detailed documentation of outsourced processes of data creation [8], remain essential steps to allow spaces of reflection, deliberation, and audit that could potentially contribute to addressing some of the social questions surrounding computer vision.

## 5. Acknowledgments

Funded by the German Federal Ministry of Education and Research (BMBF) – Nr. 16DIII13 and the International Development Research Centre (IDRC) of Canada. We dearly thank the individuals and organizations participating in this study for sharing their experiences with us.

## References

- [1] Alan Bryman. *Social Research Methods*. Oxford University Press, Oxford ; New York, 5th edition edition, 2015. 2
- [2] Kathy Charmaz. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Introducing Qualitative Methods Series. Sage Publications, London ; Thousand Oaks, Calif, 2006. 2
- [3] Juliet M. Corbin and Anselm L. Strauss. *Basics of qualitative research: techniques and procedures for developing grounded theory*. SAGE, Los Angeles, fourth edition edition, 2015. tex.ids: dns2015. 2
- [4] Mary L. Gray and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston, MA, 2019. 1
- [5] Lilly Irani. The cultural work of microwork. *New Media & Society*, 17(5):720–739, 2015. 1
- [6] Gunay Kazimzade and Milagros Miceli. Biased Priorities, Biased Outcomes: Three Recommendations for Ethics-oriented Data Annotation Practices. In *Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*, AIES '20, pages 1–7, New York, NY, USA, Feb. 2020. Association for Computing Machinery. 1
- [7] Milagros Miceli, Martin Schuessler, and Tianling Yang. Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, Oct. 2020. 1, 2
- [8] Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 161–172, Virtual Event Canada, Mar. 2021. ACM. 3

- [9] Nelson Phillips and Cynthia Hardy. *Discourse Analysis: Investigating Processes of Social Construction*. SAGE, June 2002. [2](#)
- [10] Julian Posada. The Future of Work Is Here: Toward a Comprehensive Approach to Artificial Intelligence and Labour. *Ethics of AI in Context*, 2020. [1](#), [3](#)
- [11] Paola Tubaro, Antonio A. Casilli, and Marion Coville. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, 7(1), jan 2020. [1](#)
- [12] Alex J. Wood, Mark Graham, Vili Lehdonvirta, and Isis Hjorth. Networked but Commodified: The (Dis)Embeddedness of Digital Labour in the Gig Economy. *Sociology*, 2019. [1](#)