Biased Priorities, Biased Outcomes:

Three Recommendations for Ethics-oriented Data Annotation Practices

Gunay Kazimzade*
Weizenbaum Institute for the Networked Society
Technische Universität Berlin
Berlin, Germany
gunay.kazimzade@tu-berlin.de

Milagros Miceli*
Weizenbaum Institute for the Networked Society
Technische Universität Berlin
Berlin, Germany
m.miceli@tu-berlin.de

ABSTRACT

In this paper, we analyze the relation between biased data-driven outcomes and practices of data annotation for vision models, by placing them in the context of market economy. Understanding data annotation as a sense-making process, we investigate which goals are prioritized by decision-makers throughout the annotation of datasets. Following a qualitative design, the study is based on 24 interviews with relevant actors and extensive participatory observations, including several weeks of fieldwork at two companies dedicated to data annotation for machine learning in Buenos Aires, Argentina and Sofia, Bulgaria. The prevalence of market-oriented values over socially responsible approaches is argued based on three corporate priorities that inform work practices in this field: *profit, standardization,* and *opacity*. Finally, we introduce three elements, namely *transparency, education*, and *regulations*, aiming at developing ethics-oriented practices of data annotation, that could help prevent biased outcomes.

CCS CONCEPTS

• Computing methodologies ~ Artificial intelligence • Computing methodologies ~ Machine learning • Applied computing ~ Annotation • Social and professional topics ~ Computer supported cooperative work

ACM ISBN 978-1-4503-7110-0/20/02.
https://doi.org/10.1145/3375627.3375809

KEYWORDS

Bias; data; annotation; classification; ethics; power; transparency; priority; profit

ACM Reference format:

Gunay Kazimzade, Milagros Miceli. 2020. Biased Priorities, Biased Outcomes: Three Recommendations for Ethics-oriented Data Annotation Practices. In *Proceedings of 2020 ACM AIES'20, 7-8th February, New York.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/ 3375627.3375809

1 Introduction

The dark power of data-driven systems lies in their opacity and in their ability to assign specific meaning to data with an "aura of truth, objectivity, and accuracy" [3:663]. Among the generality of algorithmic systems, machine learning algorithms can be the most deceptive ones. Because they are able to reach conclusions and make decisions on their own, that is, without having explicitly been trained for a specific output, their appearance of neutrality is all the more convincing.

The quality of data is critical for machine learning models as it holds the power to represent or exclude the population it is intended to analyze. For autonomous systems to be able to make sense of the world, humans first need to make sense of the data these systems will be trained on. This fact seems to go unnoticed quite often: despite its highly interpretative character, data-related work has many times been described as neutral, "comprising unambiguous data, and proceeding through regularized steps of analysis" [17:1]. Although the data creation structure for machine learning has been studied extensively [4,12], the ethical consequences of human-generated data annotation and its effects on data-driven outcomes have not yet attracted enough attention within the research community. To address this issue, we examine annotation practices for vision models by focusing on the values that are prioritized by decision-makers and practitioners in this field. In the pages that follow, data annotators are understood as actors interpreting data within large industrial structures. Focusing on work practices and their context, our research goal aims at understanding the logics driving companies and their impact on the performed annotations. The following research questions have guided this investigation: Which goals are prioritized by decision-makers at the data annotation stage? How do these priorities correlate with data-related bias issues?

Comprising 24 interviews, this qualitative investigation has been conducted based on fieldwork carried out in two companies dedicated to data annotation based respectively in Sofia, Bulgaria, and Buenos Aires, Argentina. In addition, managers in three further annotation companies, and engineers and data engineers in a computer vision company deploying annotated datasets were interviewed.

1 Related Work

1.1 Data and Bias

It has been argued that biases can penetrate data-driven systems at every layer of the AI pipeline, including data, design, model, and application [19]. Therefore, the quality of data is critical for the machine learning structure. Machine Learning models trained on incomplete or biased datasets can lead to discriminatory or exclusionary outcomes [5,19], especially from the recognition, classification, and decision making algorithms used in critical domains as public safety [19], credit scoring [22] or human resources management [20].

Significant academic focus [12,23–25] lies upon data annotation. Chen and Cosley [4] argue that manual annotations often yield subjective and noisy labels, as many factors can affect the quality of human-performed annotations such as unreliable annotators, poor annotation guidelines, poor category design (categories that are too broad, too narrow, or too vague), or insufficient information to make a reliable assessment. The World Economic Forum recommends building harmonized standards for data labeling and emphasizes that "All companies will benefit from greater transparency requirements around licensed datasets. This will be particularly important for startups/smaller companies who are not resourced to undergo extensive testing prior to release." [7:17]

Following this view, the current paper focuses on data annotation for computer vision models, taking into account different actors, their interaction with data, and the structures shaping those interactions. This perspective serves as a framework to investigate the correlation of decision-making in data annotation and the penetration of biases in visual models.

1.2 Data and Humans

It has been argued that data-driven systems embody specific values and interests and that these values and interests correspond with those of humans, institutions, and companies involved in their development [10,13,18]. Scholars seem to agree that a closer look into the work practices involved in the creation of data-driven systems could help untangle some of the ethical issues related to their outcomes, under consideration of the structures, conditions, and priorities shaping those practices. Even so, data-related work is still often described as a rational process of discovery, that solely aims at revealing the underlying nature of a field of inquiry, withholding the fact that data needs to be collected, analyzed, and interpreted in ways designed by humans [21].

Data transformations do not take place in a vacuum but are deeply intertwined with organizational structures and market's demands [16]. Humans collect, label, process, and analyze data in the usually invisible context of a measurement plan, where what is considered data, and how those data are measured is established. Describing this process as politics of measurements, Pine and Liboiron [21] argue that data is created by techniques of measurement that are permeated with subjective judgments and values – individual and organizational. Subjectivities are a crucial component in the complex data assemblages that frame "what is possible, desirable and expected of data" [8:24].

Building on this understanding of data as a human-influenced entity, the present paper focuses on how annotators engage with data to make sense of it, and what contextual elements are constitutive of that interaction.

1.3 Classification and Power

Analyzing classificatory practices and their relation to power is vital for understanding how technical and societal issues intrinsically relate to one another when it comes to biased outcomes: Classifications hold power to impose an arbitrary (di)vision of the world on others [2]. Both in technical and social realms, classifications are used to cluster and label varied realities so that they can be better grasped as homogenous measurable units. Previous work [13,15] has argued that data-driven systems, despite their neutrality claims, have not escaped the arbitrariness involved in the creation of taxonomies. Through classification, both humans and machines are able to sort data and make sense of it.

This perspective constitutes an essential contribution to the relation between annotation practices and exclusionary outcomes argued in the present paper. Structures that promote the apparently neutral imposition of classifications through data-driven outcomes ought to be studied if biases in machine learning are to be taken seriously.

2 Methodology

Which goals are prioritized by decision-makers at the data annotation stage? How do these priorities correlate with data-related bias issues? With our research question in mind, four sources of information have been exhaustively explored: (1) a company dedicated to data annotation located in Buenos Aires, Argentina, (2) a company dedicated to data annotation located in Sofia, Bulgaria, (3) management employees in other – larger – companies dedicated to data annotation, and (4) a computer vision company deploying annotated training sets in Berlin, Germany. In their outsourcing capacities, the studied data annotation companies base their activities in incoming projects, ordered mostly by machine learning companies developing computer vision products.

In Buenos Aires and Sofia, two qualitative data gathering methods have been deployed: participatory observation (with varying degrees of involvement) and qualitative interviewing, in the form of expert interviews and in-depth interviews. The remaining sources of information, namely the managers in other data annotation companies and the data scientists/engineers at the Berliner company, were explored through qualitative expert interviews. Phases of data collection and analysis intertwined during fieldwork and after, indicating the need to collect further data or, in time, showing that theoretical saturation had been achieved.

With an average length of nearly one hour per session, a total of 24 informants were interviewed: ten data annotators, two reviewers, four project managers, one quality assurance analysts, two data engineers, one data protection officer, one area manager, one branch manager, and two co-founders. The transcripts of those interviews were integrated with several pages of field notes from the observations conducted in Buenos Aires and Sofia, and various documents collected at both companies, containing different types of information such as specific instructions provided by clients with labeling requirements, lists of metrics for quality assurance, and impact assessments.

It is important to mention that the data used for the present paper is part of a larger dataset that was obtained for another project of our research group. Mayring's [14] qualitative content analysis was applied to analyze that data in view of the present investigation. This approach aims at interpreting the manifest and latent content of the material in their social context and field of meaning, focusing on the personal perspective of the actors [1]. This method allowed us enough flexibility to obtain valuable insights out of a dataset that had not explicitly been designed to answer our research question. Some of the topics that had been operationalized in the interview guides were indeed related to our research interest and helped us build categories for the coding of the material deductively. In addition, room for new categories was left open, so that they could be added after the exchange between recorded material and theoretical standpoints had taken place (inductive category formation). Through content analysis, we aimed at identifying patterns in the conducted interviews. Those patterns were later confronted with further sources (in this case, field notes from the conducted observations and further documents recording internal communication within the companies and with their clients). The development of coding schemes for the analysis, including categories and sub-categories, as well as the coding process itself was carried out in iterations involving cross-coding between the authors and two collaborators. We strived for interpretations that are intersubjectively comprehensible, exhaustive, and yet reflective of researchers' subjectivities. These iterative analyses led to a core set of seven coding categories and 28 sub-categories. By the end of the analysis stage, we had coded 979 statements.

3 Results and Discussion

Informed by our research questions, the above-described analysis aimed at investigating the context in which annotators make sense of data, which goals are prioritized by decision-makers, and how these priorities affect data-driven outcomes.

Klein et al. define sensemaking as "a process of framing and reframing, of fitting data into a frame that helps us filter and interpret the data" [9:119]. Considering data annotation as a praxis of sensemaking at the intersection of human subjectivity and capitalistic structures [16], our goal was not only to identify priorities but also to analyze how they are established and naturalized in work practices and industrial processes.

In the subsections that follow, we argue, on the one hand, the prevalence of three elements that inform work practices of data annotation and that are often prioritized, even when standing in opposition to more ethical approaches. The three argued priorities are profit, standardization, and opacity. They are explained and illustrated on the basis of statements from the conducted interviews. On the other hand, we further propose three elements that could help develop practices of data annotation that are more sensitive to ethical issues in artificial intelligence. The three proposed suggestions relate to the need for transparency, education, and regulations. These three ethics-oriented elements are possibly not the only ones that should be prioritized at this stage. However, they are highlighted here as they directly stand in opposition to what has been identified in this investigation, to be currently considered common practice in the field of data annotation.

3.1 How it is: three market-oriented priorities

3.1.*1* Profit

Given the for-profit character of most of the projects involving annotated data, annotation practices cannot be analyzed in a vacuum but must be put in relation to profit-oriented structures. Companies dedicated to data annotation need to position themselves in a market that demands competitive prices, fast responses, and standardized outcomes. These demands are prioritized by annotation companies and are already instilled in workers at the training stage.

Both data workers and management mention tight deadlines as one of the most problematic issues related to their work: while several of the interviewed annotators describe short time frames as the most negative aspects of their work, most of the informants in managerial positions agree that time constraints are the main reason for labeling errors. Furthermore, time constraints connected to the strive for profit are mentioned as one of the main obstacles for the implementation of processes that could help detect biased labels, such as training related to the avoidance of undesirable prejudices, transparent documentation of data transformation, and quality controls for biased labeling.

The annotation guidelines are almost always provided by the clients. Those guidelines generally aim at standardizing and optimizing labels, so they can best fit the client's product and plan to optimize revenue. Conversely, those guidelines and requirements hardly ever include instructions aiming at the avoidance of annotation-related biases. Of the six informants in managerial positions that were interviewed for this study, none recalled ever receiving a request from clients to instruct annotators in bias-related issues. Moreover, one of the quality assurance analysts interviewed in Buenos Aires described very eloquently how market logics oppose the implementation of rules against forms of sexist content, such as gender generalizations and irrelevant gender markers, on the platform of one of their clients. The informant states that the reason behind her company not having any processes in place to evaluate quality as related to the avoidance of biases in the labels lies with the clients: as long as clients do not ask (and pay) for it, it will never be implemented.

As clients prioritize profitable outcomes, approaches that could help mitigate biases do not seem to be in the scope of their priorities:

"Interviewer (I): What are the potential drivers for the implementation of the more transparent approach to documenting systems and processes?

Interview Partner (IP): If the customer demands it.

I: Is this something you have heard before, customers demanding a more...

IP: No."

Engineer with Berlin-based computer vision company.

The strive for profit seems to be prioritized over further approaches that could prevent biased outcomes. This context not only shapes internal processes, but it also influences the annotations performed.

3.1.*2* Standardization

The strive for standardized outcomes is at the core of annotation projects that are time-efficient and, thus, cost-effective. Standardized labeling practices serve both as orientation and as a constraint: on the one hand, they provide workers with a framework to perform their tasks with some certainty within well-oiled processes. On the other hand, standardization aims at constraining workers' subjectivity, thus reducing the room for questioning those classes and categories that have been instructed by clients:

"In this case we usually obey everything that they [the clients] say because, you know, their interpretations is usually the one that makes sense."

Founder of Sofia-based annotation company.

Choices regarding which platform will serve as a tool to perform the annotations and host the data are another prerogative of clients aiming at standardizing outcomes. In many cases, the client has developed their own platform, specifically tailored to the needs of their business and the desired outcomes. Technical tools constitute another form of constraint for annotators, as they shape work practices, determine what is technically possible, and limit the possible meanings that could be assigned to data:

"There was a limitation on the annotation tool that they were using. They were relying on an open-source platform that doesn't have that feature that lets you add or create predefined attributes which makes the work many times easier"

Project manager with an Iraq-based annotation company.

Quality controls are a further way of assuring the standardization of practices and the uniformity of labels. In more or less structured ways depending on the company, control instances are undertaken in iterations and aim at making sure annotations are done uniformly to fit the client's product and business plan:

"We have a quality process in order to meet the customer's requirements."

Team leader with Buenos Aires-based annotation company.

The sense-making process involved in annotation is deeply informed by standardization attempts and hierarchical impositions. Annotators rarely question instructions received at briefings. Our findings suggest that standardization provides the venue for clients (and furthermore the market) to impose their profit-oriented interests and priorities upon data. Interests and priorities that often correlate with biased outcomes.

3.1.*3* Opacity

Throughout this investigation, we were repeatedly confronted with the unfamiliarity of data annotators, experts, and decision-makers with bias-related problems, ethical implications, and the overall impact of the tasks they perform. In most of the cases, companies working on data annotation do not take responsibility for biased or discriminatory outcomes from labeled datasets. Our interview partners mostly associate bias-related issues to matters of "common sense" and clients' priorities. They emphasize the client's power of imposing their criteria on the annotation of datasets. Criteria that most of the times remain opaque due to corporate confidentiality:

"With respect to data and ethics, I must say with total honesty that we completely depend on each client and most specifically on the policies of each client"

Co-Founder of Buenos Aires-based annotation company.

Throughout the data collected for this study, evidence of the lack of documentation of the processes involved in datasets are abundant. For workers, it is at times challenging to recognize where responsibilities lie, and which factors have influenced certain decisions. Lack of transparency among the actors and layers involved in data annotation leads to gaps in quality assurance, that could otherwise help mitigate the presence of biases in datasets.

"Interviewer (I): Do you have a way of documenting the training process? I mean… a formal process.

Interview Partner (IP): No, no formal process. No. Not at all."

Engineer with Berlin-based computer vision company.

Furthermore, most of the interview partners stated not to know what the purpose of the annotations is or even what kind of products clients are developing. The informants in managerial positions relate this lack of transparency with issues of corporate confidentiality on the side of the clients. Most of them have signed non-disclosure agreements, which prevents them from accessing and sharing information:

"I: Why do they need all these pictures annotated like this? Do you know?

IP: No. I am not sure, because I never ask about this."

Project manager with Sofia-based annotation company.

Even if some of the interviewed managers were aware of biased-related hazards and possible harms resulting from data-driven systems, educating data annotators on these issues is presented as a challenging task. One that will not be undertaken anytime soon:

"It is not that the company is not aware of these things, but I think it's most now because it may be too complicated to explain to workers (...) I think it's a combination of a lot of these: the difficulties to explain and it may be the lack of curiosity or explicit curiosity on their end."

Intern with Sofia-based annotation company.

This state of things translates into workers' ignorance regarding the purpose and consequences behind the annotations they perform. Most of them have never received training on general knowledge regarding data-driven systems and machine learning, and many find it very difficult to reflect on the use and impact of their annotations, even when invited to do so during the interviews. These factors make data annotators unlikely to question those classes and categories instructed by clients no matter how biased they might be.

3.2 How it should be: three ethics-oriented priorities

3.2.*1* Transparency

The transformation of data into measurable units involves the intervention of many actors at different stages, who bring their values, subjectivities, and interests to the equation. Data collection, cleaning, and labeling are some of these transformative stages. Unfortunately, these instances of intervention and interpretation are hardly ever documented. Moreover, their iterative character and the numerous stages of control involved make it very difficult to establish where and when a given intervention has taken place, who has been involved, and, most importantly, which criteria has been followed. Under these circumstances, accountability is diluted into the many actors, layers, and iterations involved in each process, which presents a severe challenge when it comes to identifying when and how undesirable biases have crept into a model and where responsibilities lie.

In order to approach this issue and prevent discriminatory outcomes, the World Economic Forum proposes to document the provenance, creation, and use of machine learning datasets [7]. In the same vein, Gebru and colleagues [6] introduce a standardized process for documenting machine learning datasets and argue that every dataset should be accompanied with a datasheet documenting its motivation, composition, collection process, and recommended uses, among other information. This documentation should help researchers and practitioners select more appropriate datasets for their chosen tasks.

Based on the analysis of our data, we support this view and actively advocate for the implementation of transparent approaches regarding the documentation of data transformations, including information on responsibilities and criteria for decision-making. We argue that a transparent approach could contribute to assessing accountability and, in time, to mitigating biased outcomes produced by data-driven systems. Of course, the adoption of thorough documentation processes can be time-consuming and thus not quite profit-oriented. With disregard of the costs, more transparency in the development of data-driven systems should become the industry's standard if AI biases are to be fought effectively.

3.2.*2* Education

As discussed in the previous sections, biased assumptions and decisions can penetrate data annotations at different stages of the process. In many cases, these assumptions occur due to lack of training, exclusive view of the world, as well as narrow expertise on a very specific niche (image labeling, data mining, model training), which are assumed not to require knowledge on ethics. Our investigation indicates that decision-makers, managers, and data annotators are, unfortunately, not always trained to understand the global impact of unethical AI and how apparently simple work-related choices impact on society via data-driven decision making.

The Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems promoted, by the Institute of Electrical and Electronics Engineers (IEEE), highlights the urgency to "ensure every technologist is educated, trained, and empowered to prioritize ethical considerations in the design and development of autonomous and intelligent systems" [11]. The evidence collected within our investigation strongly supports this vision: training and education are fundamental to promote ethics-oriented data creation in vision models.

Training on the potential harms caused by AI and its ethical implications could help data annotators, quality assurance analysts, and managers adopt a more critical approach towards the interpretation and labeling of data. It is however critical to ground ethics-related training in the specific domain of each actor, highlighting the particular implications of their work practices as linked to the overall machine learning pipeline, its outcomes, and related societal impact.

3.2.*3* Regulations

Given the fast-paced evolution of AI being implemented in almost every sector, legal regulations have not been able to catch up with the ethical questions these developments pose. Unethical AI is generally a reflection of biased views and unethical decisions within society, in general, and within the companies and teams developing these systems, in particular. Our research findings support the argument that the lack of clear policies and ethical guidelines is one of the most significant gaps that enable the penetration of biases in data annotation.

Given this situation, we strongly advocate for clear guidelines for ethical AI to be developed at the governmental level and to be applied both in state institutions as well as in private organizations. One of our interview partners in Sofia, Bulgaria, commented on the disregard of clients for taking steps towards a more transparent approach in data-related work that could potentially mitigate unfair outcomes produced by machine learning systems. She eloquently pointed out how the lack of regulations and guidelines gives companies the freedom to approach these issues as they see fit or even to ignore them. Given this scenario, society highly depends on CEOs and managers' good will to "do the right thing." Therefore, we advocate for the urgency to prioritize societal implications as much as profit and for the importance of taking regulatory measures to accompany the pace of technological developments.

Our findings suggest that clear regulations such as the implementation and application of standards and ethical guidelines for AI could be an effective way of making sure that those companies developing data-driven systems will not only strive for profit but will also prioritize the avoidance of discriminatory outcomes. Moreover, the regulation of tools and platforms used for data annotation, incorporating standards against biases in their design, would constitute a further step toward the creation of ethical annotation practices.

4 Conclusions

Informed by insights offered by data workers and management in companies dedicated to the annotation of data for vision models, this paper has analyzed the relationship between companies' priorities and the annotations performed. We have described three market-oriented elements that are currently prioritized in this field, and that could have an impact on the quality of the annotations and therefore have a correlation with biased outcomes. These priorities are profit, standardization, and opacity. Finally, we have argued three further aspects that should be prioritized by companies and policymakers, if fighting biased labels is to be taken seriously. Those ethics-oriented aspects are transparency, education, and regulations. The research presented here is limited to the analysis of those priorities shaping data annotation and can, therefore, not yet be transferred to the broader investigations of bias and ethics in the development of AI. However, those interviews conducted with engineers at a computer vision company suggest that the framework presented in this paper could also be applied to further stages in the systems' development, such as design and model training. For future work, we plan to broaden our research question to consider further actors and more complex interplays within the machine learning pipeline.

ACKNOWLEDGMENTS

Funded by the German Federal Ministry of Education and Research (BMBF) -NR 16DII113.

We thank our colleagues at Weizenbaum Institute, Tianling Yang and Martin Schüßler for their assistance with the coding and analysis of the data. Mr. Schüßler also provided helpful comments and valuable suggestions to mature the ideas presented in this paper. Special thanks to our research group leader, Diana Serbanescu, for her continuous support of this project.

***Both authors have contributed equally to this work.**

REFERENCES

[1] Jürgen Bortz and Nicola Döring. 2001. *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (3. Aufl. 2002. 2., korr. Nachdruck 2005 edition ed.). Springer, Heidelberg.

[2] Pierre Bourdieu. 1989. Social Space and Symbolic Power. *Sociological Theory* 7, 1 (1989), 14. DOI:https://doi.org/10.2307/202060

[3] danah boyd and Kate Crawford. 2012. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15, 5 (June 2012), 662–679. DOI:https://doi.org/10.1080/1369118X.2012.678878

[4] Justin Cheng and Dan Cosley. 2013. How annotation styles influence content and preferences. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media - HT '13*, 214–218. DOI:https://doi.org/10.1145/2481492.2481519

[5] Lucas Dixon, John Li, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. 2018. Measuring and Mitigating Unintended Bias in Text Classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society - AIES '18*, 67–73. DOI:https://doi.org/10.1145/3278721.3278729

[6] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. 2018. Datasheets for Datasets. *arXiv:1803.09010 [cs]* (March 2018). Retrieved September 12, 2019 from http://arxiv.org/abs/1803.09010

[7] Global Future Council on Human Rights 2016-2018. World Economic Forum. 2018. How to Prevent Discriminatory Outcomes in Machine Learning. Retrieved from http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

[8] Rob Kitchin. 2014. *The data revolution: big data, open data, data infrastructures & their consequences*. SAGE Publications, Los Angeles, California.

[9] Gary Klein, Jennifer K. Phillips, Erica L. Rall, and Deborah A. Peluso. 2007. A data-frame theory of sensemaking. In *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 113–155.

[10] Ulrike Klinger and Jakob Svensson. 2018. The end of media logics? On algorithms and agency. *New Media & Society* 20, 12 (December 2018), 4653–4670. DOI:https://doi.org/10.1177/1461444818779750

[11] Ansgar Koene, Liz Dowthwaite, and Suchana Seth. 2018. IEEE P7003™ standard for algorithmic bias considerations: work in progress paper. In *Proceedings of the International Workshop on Software Fairness - FairWare '18*, 38–41. DOI:https://doi.org/10.1145/3194770.3194773

[12] Mani, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. 2006. Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL - ACL '06*, 753–760. DOI:https://doi.org/10.3115/1220175.1220270

[13] Steffen Mau. 2019. *The metric society: on the quantification of the social / Steffen Mau ; translated by Sharon Howe*. Polity, Cambridge ; Medford, MA.

[14] Philipp Mayring. 2014. Qualitative Content Analysis: theoretical foundation, basic procedures and software solution. (2014), 144.

[15] Milagros Miceli. 2019. AI's Symbolic Power: Classification in the Age of Automation. In *Workshop on Human-Centered Study of Data Science Work Practices. ACM CHI Conference on Human Factors in Computing Systems - CHI'19* (04.–09.05.2019), Glas- gow, UK, S. 1-4.

[16] Milagros Miceli. 2019. Data Labeling Work between Subjectivity and Standardization. A Grounded Theory Study. Unpublished master thesis, Humboldt-Universität zu Berlin.

[17] 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*, 1–15. DOI:https://doi.org/10.1145/3290605.3300356

[18] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. Retrieved August 6, 2019 from https://www.jstor.org/stable/j.ctt1pwt9w5

[19] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2016. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *SSRN Electronic Journal* (2016). DOI:https://doi.org/10.2139/ssrn.2886526

[20] Cathy O'Neil. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (01 edition ed.). Penguin, London.

[21] Kathleen H. Pine and Max Liboiron. 2015. The Politics of Measurement and Action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3147–3156. DOI:https://doi.org/10.1145/2702123.2702298

[22] Lisa Rice and Deidre Swesnik. 2013. Discriminatory Effects of Credit Scoring on Communities of Color. *Suffolk University Law Review* (2013), 32.

[23] Alexander Sorokin and David Forsyth. 2008. Utility data annotation with Amazon Mechanical Turk. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 1–8. DOI:https://doi.org/10.1109/CVPRW.2008.4562953

[24] Catherine Wah, Steve Branson, Pietro Perona, and Serge Belongie. 2011. Multiclass recognition and part localization with humans in the loop. In *2011 International Conference on Computer Vision*, 2524–2531. DOI:https://doi.org/10.1109/ICCV.2011.6126539

[25] Catherine Wah, Grant Van Horn, Steve Branson, Subhransu Maji, Pietro Perona, and Serge Belongie. 2014. Similarity Comparisons for Interactive Fine-Grained Categorization. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 859–866. DOI:https://doi.org/10.1109/CVPR.2014.115