Investigating and Mitigating Biases in Crowdsourced Data

Danula Hettiachchi
RMIT University
Melbourne, Australia

Mark Sanderson
RMIT University
Melbourne, Australia

Jorge Goncalves
The University of Melbourne
Melbourne, Australia

Simo Hosio
University of Oulu
Oulu, Finland

Gabriella Kazai
Microsoft Research
Cambridge, UK

Matthew Lease
U. of Texas at Austin & Amazon
Austin, TX, USA

Mike Schaekermann
Amazon
Toronto, Ontario, Canada

Emine Yilmaz
University College London & Amazon
London, UK

ABSTRACT

It is common practice for machine learning systems to rely on crowdsourced label data for training and evaluation. It is also well-known that biases present in the label data can induce biases in the trained models. Biases may be introduced by the mechanisms used for deciding what data should/could be labelled or by the mechanisms employed to obtain the labels. Various approaches have been proposed to detect and correct biases once the label dataset has been constructed. However, proactively reducing biases during the data labelling phase and ensuring data fairness could be more economical compared to post-processing bias mitigation approaches. In this workshop, we aim to foster discussion on ongoing research around biases in crowdsourced data and to identify future research directions to detect, quantify and mitigate biases before, during and after the labelling process such that both task requesters and crowd workers can benefit. We will explore how specific crowdsourcing workflows, worker attributes, and work practices contribute to biases in the labelled data; how to quantify and mitigate biases as part of the labelling process; and how such mitigation approaches may impact workers and the crowdsourcing ecosystem. The outcome of the workshop will include a collaborative publication of a research agenda to improve or develop novel methods relating to crowdsourcing tools, processes and work practices to address biases in crowdsourced data. We also plan to run a Crowd Bias Challenge prior to the workshop, where participants will be asked to collect labels for a given dataset while minimising potential biases.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**.

KEYWORDS

crowdsourcing, data quality, biases

CSCW '21 Companion, October 23–27, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8479-7/21/10.
https://doi.org/10.1145/3462204.3481729

ACM Reference Format:
Danula Hettiachchi, Mark Sanderson, Jorge Goncalves, Simo Hosio, Gabriella Kazai, Matthew Lease, Mike Schaekermann, and Emine Yilmaz. 2021. Investigating and Mitigating Biases in Crowdsourced Data. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion), October 23–27, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3462204.3481729

1 INTRODUCTION

While quality assurance methods are commonly used with crowdsourced data labelling, studies have shown that annotator biases often creep into labelling decisions [6, 7, 18, 31, 34]. For instance, Hube et al., [18] show that workers with strong opinions produce biased annotations, even among experienced workers. More generally, there can be different types of biases in data (e.g., population bias, behavioural bias, temporal bias) [27, 28, 35]. Mehrabi et al. [27] provide an extensive list of 23 different types of biases in data. There are several methods available to account for such biases at different stages of data preparation and model training. However, there can be economic reasons and problem-specific needs that would make certain techniques more suitable for a particular stage.

Recent work has explored promising directions that can ensure data fairness when using crowdsourcing to collect data [4, 8, 9, 18]. Instead of presenting the task to the generic worker pool, one approach is to filter the workers based on attributes that can induce bias in the labels. For example, when gathering crowd labels, it is possible to mitigate bias by using a balanced or skewed sample of workers with respect to worker demographics (e.g., Age, Gender) and the minimum wage in their country [4]. Barbosa & Chen [4] allow requesters to decide how they want to manage the worker population. For example, when collecting audio data to train a voice assistant, requesters can set the worker pool to be diverse in terms of gender, age, native language. Similarly, we can also assign questions to specific workers or worker groups taking fairness into account in addition to budget constraints and overall accuracy [9]. In addition, task presentation strategies can help in easing bias. For example, social projection, awareness reminders and personalised alerts can reduce worker bias [18].

After the data collection process, bias can also be reduced using many other approaches such as data aggregation techniques [19], systematically adding new data points to fix coverage [2], and

crowdsourced bias detection [17]. Other approaches to mitigate bias can be employed during feature engineering and model training. Prior work highlights the direct use of crowdsourced data [3], using crowds to identify perceived fairness of features [37, 38], using pre-processing methods (removing sensitive attributes, resampling the data to remove discrimination, iteratively adjust training sample weights of data from sensitive groups) [5, 20, 23] and using active learning [1]. However, techniques applied after data collection process can result in wasted effort. Furthermore, biases can be introduced as a result of many other factors such as poor task design [6], task attributes [19], due to sampling decisions the task requester makes about the data, and worker task selection preferences. Therefore, tackling biases during data collection is necessary as it may not be possible to rectify certain issues post-hoc.

Understanding and mitigating biases in crowd data is highly relevant to CSCW researchers and others who rely on crowd data for creating automated systems. In addition, researchers increasingly use crowdsourcing platforms to gather research data through survey tasks and user experiments. Several recent workshops such as Subjectivity, Ambiguity and Disagreement (SAD) in Crowdsourcing at WebConf 2019 [1], Crowd Bias workshop at HCOMP 2018 [2], Crowd Science at NeurIPS 2020 [3], and Data Excellence at HCOMP 2020 [4] have explored related topics with a particular focus on data, evaluation and applications. While continuing the broader discussion on biases in crowd data, in this workshop, we aim to focus on crowd workers and the crowdsourcing process. Therefore, we anticipate this workshop and its outcomes will be relevant and important to the broader CSCW community.

2 WORKSHOP

2.1 Goals and Themes

Through this workshop, we aim to foster discussion on ongoing work around biases in crowd data, provide a central platform to revisit the current research, and identify future research directions that are beneficial to both task requesters and crowd workers. We have identified the following four workshop themes that will focus on understanding, quantifying and mitigating biases in crowd data while exploring their impact on crowd workers.

1. *Understanding how annotator attributes contribute to biases*: Research on crowd work has often focused on task accuracy whereas other factors such as biases in data have received limited attention. We are interested in reviewing existing approaches and discussing ongoing work that helps us better understand annotation attributes contributing to biases.

2. *Quantifying bias in annotated data*: An important step towards bias mitigation is detecting such biases and measuring the extent of biases in data. We seek to discuss different methods, metrics and challenges in quantifying biases, particularly in crowdsourced data. Further, we are interested in ways of comparing biases across different samples and investigating if specific biases are task-specific or task-independent.

3. *Novel approaches to mitigate crowd bias*: We plan to explore novel methods that aim to reduce biases in crowd annotation in particular. Current approaches range from worker pre-selection, improving task presentation and dynamic task assignment. We seek to discuss shortcomings and limitations of existing and ongoing approaches and ideate future directions.

4. *Impact on crowd workers*: We want to explore how bias identification and mitigation strategies can impact the actual workers, positively or negatively. For example, workers in certain groups may face increased competition and lack of task availability. Collecting worker attributes and profiling could raise ethical concerns.

2.2 Structure and Activities

We have planned the workshop structure with a specific focus on virtual participation. Considering the potential spread of participants across multiple time zones, we plan on running a 5-hour synchronous workshop and an optional workshop challenge as a pre-workshop activity. In addition, we will invite workshop participants to submit position papers relating to workshop themes.

2.3 Pre-workshop Activities

Crowd Bias Challenge: To maximise the networking opportunities and participant engagement, we plan to introduce a workshop challenge (e.g., TREC Challenges) where participants will gather a crowdsourced dataset for a given problem while minimising potential biases in data. The challenge will start three weeks before the conference, and participants will be invited to attempt the challenge as individuals or in groups of up to 4 members. We will use a milestone and result-oriented leaderboard to motivate the teams.

The challenge will introduce a problem where biases in data are problematic (e.g., crowd judgements on content moderation). We will be providing the seed dataset for the participants to gather labels. To set up the crowd task and gather data, the teams will be provided with access to a crowdsourcing platform with a limited amount of credits (e.g., Amazon MTurk). The teams will then submit both original crowd labels and the aggregated result that they come up with. The results will be evaluated using pre-specified ground truth data and content categories that are susceptible to biases. For instance, we will consider the variation in accuracy across content categories in addition to overall task accuracy.

Position Papers: We will invite the participants to submit 2-page position papers on previous or ongoing research work on biases in crowd data.

2.4 Synchronous Workshop

The synchronous workshop during the conference will include the following sessions. We will also appropriately incorporate breaks and social activities into the final schedule.

Introduction (1 hour): We will provide a brief introduction to the workshop outlining planned activities, goals and themes of the workshop. We also plan to run a quick ice-breaking round to get to know the participants.

Position Paper Discussion (1.5 hour): In the synchronous session, participants will share their position papers with the audience. We plan on allocating time for selected papers, and the presentations

[1] https://sadworkshop.wordpress.com/
[2] https://sites.google.com/view/crowdbias
[3] https://research.yandex.com/workshops/crowd/neurips-2020
[4] http://eval.how/dew2020/

will be organised under workshop themes. Each presenter will get 3 minutes to present their work followed by 7 minutes of questions from the audience. We want to prioritise the opportunity for authors to obtain feedback from the audience. We will also share the position papers with participants prior to the workshop and encourage them to read them beforehand.

Crowd Bias Challenge Recap (1 hour): The challenge outcomes will be presented and discussed during the workshop. We will invite the leading three teams to discuss their solutions briefly.

Blue sky session - beyond the crowd (1 hour): This ideation session will be facilitated by an invited expert outside the crowdsourcing research domain to explore challenging future research directions. We will probe participants to discuss more broad research questions around biases and crowd data. For example, how can we scale crowdsourcing platforms to wider population groups while preserving bias considerations? We plan to conduct small group discussions under four themes.

Closing (30 mins). We will use this session for summarising the workshop output and obtaining feedback from participants regarding possible future activities. We will also facilitate follow-up conversations after formally concluding the workshop.

2.5 Post Workshop Activities

We will maintain the slack workspace to facilitate follow up conversations. We also plan to publish the outcome of the crowd bias challenge and anticipate that other successful participant teams will extend their work to similar outputs. With permission of participants, we hope to publish workshop proceedings through CEUR (http://ceurws.org). In addition, based on workshop discussions, we intend to document a list of biases in crowdsourced data, their sources and suggested methods to mitigate them during or after labelling.

2.6 Virtual Setup and Participants

We plan to recruit 20-30 participants for the workshop. Participants will be required to either submit a position paper or take part in the workshop challenge. Our workshop will align with the interests of researchers and practitioners in crowdsourcing, CSCW, HCI and IR who explore biases in crowd data. The workshop will also appeal to individuals who use crowdsourcing for applications in broader areas of machine learning, social science and data science. We will promote our workshop, challenge and the position paper call via online mailing lists, social media and forum. The workshop will also be actively promoted through research groups, and centres organisers are affiliated with (e.g., multidisciplinary ARC Centre of Excellence on Automated Decision Making and Society [5]). We plan to utilise the following tools to deliver the virtual workshop.

Workshop Website: Workshop website will publish all public information including the call for position papers. *Slack Workspace:* We will setup a slack workspace dedicated to the workshop to enable communication among participants, particularly during the pre-workshop activities and to support followup conversations. We will also use the workspace to share the crowd bias challenge material and accepted position papers prior to the workshop. *Virtual Workshop Video Conferencing Platform*: We will use Zoom or Microsoft Teams to run the synchronous session of the workshop. Both platforms have features that allow us to create breakout rooms, share content and secure the session.

In addition, we plan to incorporate regular interactive polls to get continuous feedback from the participants and maximise the engagement.

[5] https://www.admscentre.org.au/

3 ORGANISERS

Our team consists of scholars and industry leaders working in and across CSCW, HCI, IR and Crowdsourcing. In addition to a strong record of being part of conference organising activities, the team also have prior experience with related workshops [24, 29] and running challenges [25, 30].

Danula Hettiachchi is a Research Fellow at the ARC Centre of Excellence on Automated Decision Making and Society and RMIT School of Computing where he research user biases when interacting with automated systems. His doctoral research examined task assignment in crowdsourcing [13, 14].

Mark Sanderson is a Professor of information retrieval (i.e. search engines). He has published extensively in the areas of evaluation of search engines [30], user interaction with search, and conversational searching systems. Mark is a Chief Investigator on the $32 million Australian Government Centre of Excellence, Automated Decision Making and Society (ADMS). He is a visiting professor at the National Institute of Informatics in Tokyo.

Jorge Goncalves is a Senior Lecturer in the School of Computing and Information Systems at the University of Melbourne. He has conducted extensive research on improving crowd data quality [10, 14, 15], and bringing crowdsourcing beyond the desktop by using ubiquitous technologies [11, 12]. He has also served as Workshops Co-Chair for CHI'19 and CHI'20.

Simo Hosio is an Associate Professor and the Principal Investigator of the Crowd Computing Research Group at the Center for Ubiquitous Computing, University of oulu, Finland. His research focuses on social computing, crowd-powered solutions for digital health [16], and crowdsourced creativity.

Gabriella Kazai is a Principal Applied Scientist at Microsoft, focusing on offline evaluation, crowdsourcing, and metric development for various search scenarios, including organic web search, news, autosuggestions, and covering aspects from relevance to source credibility. She co-organised a number of evaluation initiatives, including the TREC crowdsourcing track [25, 36] and INEX, and currently serves on the Steering Committee of the AAAI Human Computation and Crowdsourcing (HCOMP) conference and as PC chair for SIGIR 2022. Her research interests include information retrieval evaluation [30], human computation, gamification, recommender systems, and information seeking behaviour.

Matthew Lease is an Associate Professor in the School of Information at UT Austin and an Amazon Scholar in Amazon's Human-in-the-Loop services. He has conducted extensive research in crowdsourcing and human computation for a decade [13, 22, 24, 26, 30, 36] and currently serves as Steering Committee co-chair of the AAAI Human Computation and Crowdsourcing (HCOMP) conference.

Mike Schaekermann [13, 32, 33] is an Applied Scientist in Amazon's Human-in-the-Loop services. His dissertation work in crowdsourcing and human computation was recently honoured with the 2020 *Distinguished Dissertation Award* from CS-Can|Info-Can, Canada's key computer science professional society.

Emine Yilmaz [21, 24, 26, 35] is a Professor and Turing Fellow at University College London, Department of Computer Science. She also works as an Amazon Scholar with Amazon Alexa Shopping. She has been working on modelling and evaluating annotator quality, user modelling and evaluating bias in information retrieval systems.

REFERENCES

[1] Hadis Anahideh, Abolfazl Asudeh, and Saravanan Thirumuruganathan. 2020. *Fair active learning*. Vol. 1. ACM. arXiv:2001.01796

[2] Abolfazl Asudeh, Zhongjun Jin, and H. V. Jagadish. 2019. Assessing and remedying coverage for a given dataset. *Proceedings - International Conference on Data Engineering* 2019-April (2019), 554–565. https://doi.org/10.1109/ICDE.2019.00056

[3] Agathe Balayn, Panagiotis Mavridis, Alessandro Bozzon, Benjamin Timmermans, and Zoltán Szlávik. 2018. Characterising and mitigating aggregation-bias in crowdsourced toxicity annotations. *CEUR Workshop Proceedings* 2276 (2018), 67–71.

[4] Natã M. Barbosa and Monchu Chen. 2019. Rehumanized crowdsourcing: A labeling framework addressing bias and ethics in machine learning. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–12. https://doi.org/10.1145/3290605.3300773

[5] Flavio P. Calmon, Dennis Wei, Karthikeyan Natesan Ramamurthy, and Kush R. Varshney. 2017. Optimized data pre-processing for discrimination prevention. *arXiv* Nips (2017).

[6] Carsten Eickhoff. 2018. Cognitive Biases in Crowdsourcing. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*. ACM, USA, 162–170. https://doi.org/10.1145/3159652.3159654

[7] Mor Geva, Yoav Goldberg, and Jonathan Berant. 2019. Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (2019), 1161–1166.

[8] Bhavya Ghai, Q. Vera Liao, Yunfeng Zhang, and Klaus Mueller. 2020. Measuring Social Biases of Crowd Workers using Counterfactual Queries. In *Workshop on Fair and Responsible AI at ACM CHI 2020*.

[9] Naman Goel and Boi Faltings. 2019. Crowdsourcing with Fairness, Diversity and Budget Constraints. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. ACM, USA, 297–304. https://doi.org/10.1145/3306618.3314282

[10] Jorge Goncalves, Simo Hosio, Jakob Rogstadius, Evangelos Karapanos, and Vassilis Kostakos. 2015. Motivating participation and improving quality of contribution in ubiquitous crowdsourcing. *Computer Networks* 90 (2015), 34–48. https://doi.org/10.1016/j.comnet.2015.07.002 Crowdsourcing.

[11] Jorge Goncalves, Simo Hosio, Niels van Berkel, Furqan Ahmed, and Vassilis Kostakos. 2017. CrowdPickUp: Crowdsourcing Task Pickup in the Wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 51 (Sept. 2017), 22 pages. https://doi.org/10.1145/3130916

[12] Jorge Goncalves, Hannu Kukka, Iván Sánchez, and Vassilis Kostakos. 2016. Crowdsourcing Queue Estimations in Situ. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, USA, 1040–1051. https://doi.org/10.1145/2818048.2819997

[13] Danula Hettiachchi, Mike Schaekermann, Tristan J. McKinney, and Matthew Lease. 2021. The Challenge of Variable Effort Crowdsourcing and How Visible Gold Can Help. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021). arXiv 2105.09457.

[14] Danula Hettiachchi, Niels van Berkel, Vassilis Kostakos, and Jorge Goncalves. 2020. CrowdCog: A Cognitive Skill based System for Heterogeneous Task Assignment and Recommendation in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (oct 2020), 1–22. https://doi.org/10.1145/3415181

[15] Simo Hosio, Jorge Goncalves, Vili Lehdonvirta, Denzil Ferreira, and Vassilis Kostakos. 2014. Situated Crowdsourcing Using a Market Model. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, USA, 55–64.

[16] Simo Johannes Hosio, Niels van Berkel, Jonas Oppenlaender, and Jorge Goncalves. 2020. Crowdsourcing Personalized Weight Loss Diets. *Computer* 53, 1 (2020), 63–71. https://doi.org/10.1109/MC.2019.2902542

[17] Xiao Hu, Haobo Wang, Anirudh Vegesana, Somesh Dube, Kaiwen Yu, Gore Kao, Shuo-Han Chen, Yung-Hsiang Lu, George K. Thiruvathukal, and Ming Yin. 2020. Crowdsourcing Detection of Sampling Biases in Image Datasets. In *Proceedings of The Web Conference 2020 (WWW '20)*. ACM, USA, 2955–2961. https://doi.org/10.1145/3366423.3380063

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

[19] Ece Kamar, Ashish Kapoo, and Eric Horvitz. 2015. Identifying and Accounting for Task-Dependent Bias in Crowdsourcing. *Proceedings, The Third AAAI Conference on Human Computation and Crowdsourcing (HCOMP-15)* (2015), 92–101.

[20] Faisal Kamiran and Toon Calders. 2012. *Data preprocessing techniques for classification without discrimination*. Vol. 33. 1–33 pages. https://doi.org/10.1007/s10115-011-0463-8

[21] Ömer Kırnap, Fernando Diaz, Asia Biega, Michael Ekstrand, Ben Carterette, and Emine Yilmaz. 2021. Estimation of Fair Ranking Metrics with Incomplete Judgments. In *Proceedings of the Web Conference 2021 (WWW '21)*. ACM, USA, 1065–1075.

[22] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1301–1318.

[23] Emmanouil Krasanakis, Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2018. Adaptive sensitive reweighting to mitigate bias in fairness-aware classification. *Proceedings of the Web Conference 2018* 2 (2018), 853–862. https://doi.org/10.1145/3178876.3186133

[24] Matthew Lease, Vitor Carvalho, and Emine Yilmaz (Eds.). 2010. *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*. Online, Geneva, Switzerland. http://ir.ischool.utexas.edu/cse2010/materials/CSE2010-Proceedings.pdf

[25] Matthew Lease and Gabriella Kazai. 2011. Overview of the trec 2011 crowdsourcing track. In *Proceedings of the text retrieval conference (TREC)*.

[26] Matthew Lease and Emine Yilmaz. 2013. Crowdsourcing for Information Retrieval: Introduction to the Special Issue. *Information Retrieval* 16, 2 (April 2013), 91–100.

[27] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2019. A survey on bias and fairness in machine learning. arXiv:1908.09635

[28] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data* 2 (2019). https://doi.org/10.3389/fdata.2019.00013

[29] Jonas Oppenlaender, Maximilian Mackeprang, Abderrahmane Khiat, Maja Vuković, Jorge Goncalves, and Simo Hosio. 2019. DC2S2: Designing Crowd-Powered Creativity Support Systems. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, USA, 1–8. https://doi.org/10.1145/3290607.3299027

[30] Adam Roegiest et al. 2019. FACTS-IR: Fairness, Accountability, Confidentiality, Transparency, and Safety in Information Retrieval. *SIGIR Forum* 53, 2 (2019).

[31] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The risk of racial bias in hate speech detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics*. 1668–1678.

[32] Mike Schaekermann, Carrie J Cai, Abigail E Huang, and Rory Sayres. 2020. Expert Discussions Improve Comprehension of Difficult Cases in Medical Image Assessment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems - CHI '20*. ACM, USA. https://doi.org/10.1145/3313831.3376290

[33] Mike Schaekermann, Joslin Goh, Kate Larson, and Edith Law. 2018. Resolvable vs. Irresolvable Disagreement: A Study on Worker Deliberation in Crowd Work. In *Proceedings of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2018)*, Vol. 2. USA, 1–19. https://doi.org/10.1145/3274423

[34] Shilad Sen, Margaret E Giesel, Rebecca Gold, Benjamin Hillmann, Matt Lesicko, Samuel Naden, Jesse Russell, Zixiao Wang, and Brent Hecht. 2015. Turkers, Scholars,"Arafat" and "Peace" Cultural Communities and Algorithmic Gold Standards. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 826–838.

[35] Milad Shokouhi, Ryen White, and Emine Yilmaz. 2015. Anchoring and Adjustment in Relevance Estimation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. ACM, USA, 963–966.

[36] Mark Smucker, Gabriella Kazai, and Matthew Lease. 2013. Overview of the TREC 2012 Crowdsourcing Track. In *Proceedings of the 21st NIST Text Retrieval Conference (TREC)*.

[37] Niels Van Berkel, Jorge Goncalves, Danula Hettiachchi, Senuri Wijenayake, Ryan M. Kelly, and Vassilis Kostakos. 2019. Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019). https://doi.org/10.1145/3359130

[38] Niels van Berkel, Jorge Goncalves, Daniel Russo, Simo Hosio, and Mikael B. Skov. 2021. Effect of Information Presentation on Fairness Perceptions of Machine Learning Predictors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, USA, Article 245, 13 pages. https://doi.org/10.1145/3411764.3445365