Responsibly Training Foundation Models: Actualizing Ethical Principles for Curating Large-Scale Training Datasets in the Era of Massive AI Models

MORGAN KLAUS SCHEUERMAN, Sony AI, USA
DORA ZHAO, Stanford University, USA
JERONE T. A. ANDREWS, Sony AI, UK
ABEBA BIRHANE, Trinity College Dublin, Ireland
Q. VERA LIAO, University of Michigan, Ann Arbor, USA
GEORGIA PANAGIOTIDOU, King's College London, UK
POOJA CHITRE, Arizona State University, USA
KATHLEEN PINE, Arizona State University, USA
SHAWN WALKER, Arizona State University, USA
JIEYU ZHAO, University of Southern California, USA
ALICE XIANG, Sony AI, USA

AI technologies have become ubiquitous, influencing domains from healthcare to finance and permeating our daily lives. Concerns about the values underlying the creation and use of datasets to develop AI technologies are growing. Current dataset practices often disregard critical ethical issues, despite the fact that data represents and impacts real people. While progress has been made in establishing best practices for curating smaller datasets in a more ethical fashion, the unprecedented scale of training data in the era foundation models presents unique hurdles for which AI researchers and practitioners must now face. This workshop aims to unite interdisciplinary researchers and practitioners in an effort to identify the challenges unique to curating datasets for large-scale foundation models—and then begin to ideate best practices for tackling those challenges. Drawing from CSCW's tradition of interdisciplinary exchange, our aim is to cultivate a diverse community of researchers and practitioners interested in defining the future of ethical responsibility in the *composition*, *process*, and *release* of large-scale datasets for foundation model training. We will disseminate the outcomes of this workshop to the HCI community and beyond by developing a conceptual framework of both the challenges and potential solutions associated specifically with curating datasets for foundation models.

CCS Concepts: • **Computing methodologies → Artificial intelligence**; • **Human-centered computing → Human computer interaction (HCI)**; • **Social and professional topics → Socio-technical systems**.

Additional Key Words and Phrases: Fairness, ethics, responsible AI, foundation models, generative AI, datasets, machine learning, responsible artificial intelligence, human-centric artificial intelligence, algorithmic bias, values in design, work practice

ACM Reference Format:
Morgan Klaus Scheuerman, Dora Zhao, Jerone T. A. Andrews, Abeba Birhane, Q. Vera Liao, Georgia Panagiotidou, Pooja Chitre, Kathleen Pine, Shawn Walker, Jieyu Zhao, and Alice Xiang. 2018. Responsibly Training Foundation Models: Actualizing Ethical

Principles for Curating Large-Scale Training Datasets in the Era of Massive AI Models. In . ACM, New York, NY, USA, 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

AI is increasingly integrated into various practical domains. Its versatility is evident in its use across a range of domains, from healthcare [19, 22] to financial services [20, 56], playing a pivotal role in our daily lives. This versatility has never been more apparent than with the proliferation of *foundation models*, large-scale models trained on a broad and enormous array of data with the goal of being applied downstream to a wide variety of tasks [17].

There have been widespread concerns regarding the datasets used to develop these AI systems [11, 13, 18, 58, 68]. In particular, concerns about AI datasets encompass issues related to the *composition* of data in the dataset [11, 13, 18, 24, 24, 35, 42, 68, 96], the *process* of collecting and labeling it [5, 6, 11, 25, 40, 41, 47, 60, 71, 72, 75], and the *release* of the data for broader use [9, 31, 77]. Nonetheless, current practices in dataset curation for AI often prioritize dataset size and utility, overlooking critical issues like fairness, privacy, and sustainability—despite the fact that "most data represent or impact people" [98]. As datasets continue to scale in the era of foundation models, attending to the ethical implications surrounding the compositions, processes, and release of massive datasets becomes uniquely challenging.

In this workshop, we aim to: (1) define the ethical principles that should apply to the composition, process, and release of large-scale human-centric[1] training datasets; (2) address the challenges that stand in the way of enacting ethical principles given the size and technical needs of foundation models, specifically; and (3) ideate opportunities for overcoming those challenges to define best practices for curating large-scale training datasets responsibly and imagine potential solutions such as tooling (e.g. [36, 54]), policies (e.g., [1, 65]), and frameworks (e.g., [34, 48]).

With the continuous growth in interest in the societal implications of foundation models within the HCI community broadly (e.g., [21, 53, 85]) and CSCW specifically (e.g., [7, 50, 55]), we believe this workshop will garner substantial interest among CSCW participants and, as a result, garner engaged, diverse, and fruitful insights and future collaborations. Given the interdisciplinary nature of the CSCW community, we believe that CSCW is the best venue to kick off this workshop series as the development and release of foundation models continues to rapidly grow, largely unchecked.

2 ETHICAL PRINCIPLES FOR LARGE-SCALE DATASET CURATION

As AI has become globally ubiquitous, so too have the harms caused by AI deployments [32, 33, 45, 74, 87]. As a reaction to these harms, numerous scholars have sought to define *ethical principles* aimed at guiding AI's development and deployment. For example, the ethical principle of "beneficence" is focused on providing benefit to others [23]. Yet the beliefs underlying the concept of ethical principles are vast, difficult to define, and often inconsistent [34, 62, 95], leading some scholars to besmirch the idea of AI ethics altogether [38, 73, 81]. Largely, the root of the issue with AI ethics is the gap between principle and action [63], especially given that a single ethical principle may result in numerous outcomes in practice [15]. Many practitioners trying to enact ethical principles are unsure how to proceed, given the different priorities of individual actors and larger organizations [2, 44, 49]. How to enforce responsibility for enacting ethics is an open challenge.

In particular, ethical principles are often most impactful when constrained to specific model tasks (e.g., [90, 93]) or data types (e.g., [16, 59]). However, in the case of foundation models, the data collected is so vast and broad that it is intentionally not constrained to any specific task or type. While we agree with prior work that defining ethical principles and responsibly putting them into action is difficult, the curation of large-scale datasets for training foundation models

[1] Data that centers human faces, bodies, and cultural concepts [77].

is only increasing. To guide specific practices for diminishing the harms of foundation model datasets, it is necessary to define what those harms are and what principles should guide mitigating them downstream. Given the vast array of possible principles, we plan to scope workshop discussions to the five broad principles Jobin et al. found underlying the institutional documentation of AI ethics across the globe: transparency, justice and fairness, non-maleficence, responsibility, and privacy [51].

Workshop Objective 1: Define the underlying qualities associated with five ethical principles (transparency, justice and fairness, non-maleficence, responsibility, and privacy) by which large-scale dataset curation should be guided.

3 CHALLENGES TO ETHICAL LARGE-SCALE DATASET CURATION

Responsible dataset curation is not simple or straightforward. Even in instances where researchers have attempted to collect fair or otherwise ethical datasets, they have been found to incidentally violate the expectations of other researchers or the public [30, 91, 97]. Existing datasets have violated data subject consent [5, 11, 71, 77], infringed on copyright [28, 52, 76], exploited data workers [8, 37, 89], contained poor demographic distribution [14, 43, 94] and offensive image labels [10, 12, 84], and accidentally included illegal content [83]. These concerns span different stages of the dataset lifecycle, including their composition, the processes underlying their creation, and their release for academic and commercial uses.

Yet collecting datasets responsibly is extremely challenging, even when those datasets are relatively small in scale. Recent work from Zhao et al. uncovered extensive challenges that dataset curators face when trying to enact ethics throughout the dataset curation lifecycle [95]. The kinds of challenges which exist when creating fair evaluation datasets may be much more complex for large-scale training datasets—or even entirely different.

Take ImageNet, once the standard for computer vision (CV) model training, which has around 14 million images [27]. Now, ImageNet is considered too small for training foundation models. Datasets like LAION-5B, with its over five billion images, have become the new standard [78]. How would the challenges associated with a lack of resources impact approaches to dataset ethics for five billion human images? Using the ethical principles for large-scale datasets fleshed out by workshop participants (Workshop Objective 1), we then plan to discuss the challenges to achieving those principles given the massive scope and scale of large-scale datasets for training foundation models.

Workshop Objective 2: Identify the challenges specific to curating ethical large-scale training datasets.

4 OPPORTUNITIES FOR RESPONSIBLE LARGE-SCALE DATASET CURATION

While there are certain to be outstanding and thorny challenges to responsibly curating large-scale datasets that adhere to ethical principles, it is still necessary to attempt to overcome them in our shared goal to create ethical foundation models downstream. Already, scholars have improved numerous suggestions for positively improving dataset qualities like increasing data diversity [64, 67], obtaining data subject consent [57, 61, 88], providing fair wages to data workers [3, 29, 80], limiting dataset use [70], engaging stakeholders in data taxonomy design [66, 69, 79], and creating transparent documentation [26, 46, 92]. Andrews et al. provided a comprehensive framework of ethical considerations for responsible dataset curation throughout the development lifecycle, illuminating idealistic data curation approaches for smaller scale evaluation datasets [4].

Yet, many of these ethical approaches may need to be reconsidered and redefined for the scale of foundation model data. There may also be need for entirely new approaches which have yet to be considered, especially around issues like data instability and recency, transparency tools for parsing massive unstructured datasets, and environmental stability for both collection and use [39, 82, 86].

Having identified the challenges (Workshop Objective 2) to responsibly enacting the ethical principles (Workshop Objective 3), the final aim of the workshop is to begin initial ideation as an interdisciplinary community of research and practice to *actualizing ethics responsibly* for large-scale dataset curation. For each challenge associated with dataset composition, process, and release, workshop participants will ideate approaches to responsibly actualizing ethical principles to overcoming them.

Workshop Objective 3: Ideate potential approaches for responsibly curating large-scale training datasets that adhere to ethical principles.

5 WORKSHOP GOALS

As every industry scrambles to build and adopt foundation models, it is imperative that we as a community identify and define ethical standards to uphold when curating the massive datasets underlying such models. This workshop aims to gather interdisciplinary researchers and practitioners who are interested in addressing the challenges associated with creating, managing, and using data responsibly for training large-scale foundation models, like those underlying ChatGPT, BERT, and DALL-E. Given CSCW's's rich history of interdisciplinary discourse, we plan to engage participants from a diverse range of communities and backgrounds and encourage the sharing of ideas across topics and domains. The workshop will gather interdisciplinary researchers and practitioners interest in the use of human-centric data for training foundation models, including generative AI, LLMs, and other large-scale AI tasks.

In this workshop, we aim to address: (1) the qualities underlying *ethical principles* as they apply to large-scale datasets used to train foundation models; (2) the *challenges* specifically associated with responsibly curating datasets for large-scale foundation models that adhere to desired ethical principles; and (3) the potential *opportunities* for mitigating those challenges and promoting responsible dataset curation in an era of large-scale foundation model training. We aim to address key questions during the workshop, such as:

- How do **existing challenges** to enacting ethics via responsible data curation apply to large-scale foundation models?
- What are the **unique challenges** specific to actualizing ethics when curating datasets for large-scale foundation models?
- What are the **cultural, technical, social, legal, and environmental factors** that should be prioritized when defining ethical principles for dataset curation?
- What **existing ethical principles and approaches** to responsible data curation can be applied to large-scale foundation model data?
- How can we **assess the effectiveness** of ethical principles for large-scale foundation model training datasets?
- What **challenges exist for different parts of the dataset curation lifecycle**?
- What **opportunities are specific to the different parts of the dataset curation lifecycle**?
- How might we **develop regulation** specific to upholding ethical principles in training data for large-scale foundation models?
- How do we **assign responsibility** for large-scale ethical dataset curation?
- What practices and conditions should be implemented to ensure **ethical labor standards in data collection and annotation** processes?

- How can we **encourage the adoption** of ethical dataset practices within academic and industry settings?
- What role can **interdisciplinary and cross-domain collaboration** play in developing ethical principles and enacting responsible dataset curation?
- What strategies can be employed to **engage with underrepresented communities** and involve them in the dataset creation process?
- What are the considerations and potential **consequences of specific dataset collection methodologies**?
- What **tools, policies, and processes** can enable more responsible large-scale training dataset curation?

6 WORKSHOP LOGISTICS

Pre-workshop plans: Our pre-workshop plans will focus on: (1) *advertising* the workshop so that we receive strong and diverse submissions; (2) *building community* among workshop participants; and (3) *knowledge sharing* prior to the workshop.

First, to advertise the workshop to an interdisciplinary audience, we will circulate calls for participation using both mailing lists and social media. The organizing committee has a strong network across CSCW, HCI, AI, FATE, and social science communities. We will use social media websites like X/Twitter, Bluesky, LinkedIn, and Facebook groups (e.g., ACM SIGCHI) to reach familiar and new audiences. We will also distribute calls for participation via listservs and communication channels including those related to HCI (CSCW-SIG, CSCW-ALL, CHI-announcements, CHI-resources), algorithmic fairness (e.g., FAccT, AIES, Algo Audit Network), machine learning (e.g., MLCommons, Data-centric Machine Learning Research, Natural Language Processing Data Community), and social computing (e.g., 4S, AIR-L, Labortech, AOIR). We will also circulate the call to personal contacts who might be interested in the workshop. To promote diversity, equity, and inclusion, we will promote participation from historically marginalized and underrepresented groups.

Second, to build community, we will create a Slack space or Discord channel with details about the workshop, calls for participation, and instructions for submissions. This online space will be used both before and during the workshop. Before the workshop, it will be used to facilitate planning and introductions between participants before the workshop.

Finally, we plan to share reading materials focused on ethical principles and challenges to dataset curation with participants on this platform approximately *two weeks prior to the workshop day*. We will also prompt to introduce themselves and their personal goals for the workshop. We will include workshop day instructions and questions to think over as well.

Workshop mode:

Our workshop will also be largely design and discussion focused. CSCW 2025 will be a primarily in-person event with limited support for virtual attendance. Thus, we plan to host our workshop as an in-person event, to be able to facilitate the most successful workshop community possible. To best conduct this in-person workshop, we will require A/V support to facilitate participant lighting talks and final presentations at the end of the session. To facilitate group sessions, we will require a large room with six to eight large tables. The tables should be far enough apart from one another to allow teams to work privately together during group sessions. Attendees will be asked to bring personal devices so that they can work using Miro or other virtual brainstorming boards. We will also ask participants to take notes about their discussions using Google docs.

Start	End	Duration	Session
			Before the Workshop
-	-	2 weeks	Participants introduce themselves in the Discord channel
			Day of the Workshop (9:00–16:00 JST)
9:00	9:30	30 min	Welcome and Opening Remarks
9:30	10:30	60 min	Participant Lightning Talks
10:30	11:00	30 min	Coffee Break #1
11:30	12:30	60 min	Group Session # 1: Ethical Principles for Dataset Curation
12:30	1:30	60 min	Lunch Break
13:30	14:30	60 min	Group Session #2: Challenges to Ethical Dataset Curation
14:30	14:45	15 min	Coffee Break #2
14:45	15:45	60 min	Group Session #3: Opportunities for Ethical Dataset Curation
15:45	16:00	15 min	Coffee Break #3
16:00	16:45	45 min	Group Session #4: Framework Writing
16:45	17:45	60 min	Group Share
17:45	18:00	15 min	Closing and final remarks
			Optional: Post-Workshop Dinner

Table 1. Tentative schedule of workshop activities, including asynchronous activities prior to the workshop day.

7 WORKSHOP DAY

The workshop will be organized as a one-day in-person event tentatively taking place from 9:00 to 18:00 in Bergen, Norway (see Table 1 for the proposed schedule). We expect an attendance of around 15-35 total participants. In case of larger numbers, the lightning talk sessions will be organized in breakout groups that will be created based on themes from the position papers. However, we plan to cap attendance to 35 to facilitate more engaged conversations in the limited time of the workshop.

For the group sessions aimed at attending to the three workshop goals, we will establish groups that will persist throughout the workshop to ensure continuity of topic understanding for the sake of framework writing and group sharing at the end of the workshop day. Groups will be formed around the three areas of ethical concern highlighted in the Introduction above: (1) composition; (2) process; and (3) release. Each group session will correspond to each workshop objective introduced in Sections 2, 3, and 4 above, respectively. The culmination of each group activity will be clustered in Group Session #4, where groups will be asked to design a draft framework for their respective area of ethical concern. Groups will then be asked to share high-level takeaways of these preliminary frameworks.

We also plan to host an optional post-workshop dinner so that workshop participants can further exchange ideas and network with one another, including those they did not get a chance to interact with as much during the workshop.

8 POST-WORKSHOP PLANS

After the conclusion of the workshop, the organizers will provide a brief summary of the workshop on the website (https://responsiblefmdata.github.io/) and Discord channel. Position papers will also be published on the website and as collated proceedings on ArXiv, with author permission. Beyond the inaugural workshop at CSCW 2025, we plan to host a series of workshops focused on ideating and refining best practices for ethically curating large-scale training datasets with different scholarly communities, including: at machine learning conferences (NeurIPS), design conferences (DIS), and fairness conferences (FAccT and/or AIES). We plan to use the Discord for future workshops in the workshop series. This will facilitate continued participation and community building as the workshop series develops across conference communities. We will then collate and analyze our longitudinal takeaways in the form of an article or white paper so that the broader community can learn from the shared knowledge of the workshop community. Interested workshop participants will be invited to contribute to this article or white paper. We also hope to publish a special issue of a journal (e.g., TOCHI) focused on ethical principles, challenges, and opportunities for dataset curation. This special issue will serve as a platform for workshop participants to either expand upon their position papers, refining them for potential publication after the workshop, or to submit any other pertinent work they may have developed.

9 CALL FOR PARTICIPATION

Call for participation: The workshop website (https://responsiblefmdata.github.io/) will host information about the workshop and its goals, as well as instructions to apply to attend. A strong emphasis will be placed on promoting a broad range of viewpoints on fairness in ML data across a variety of domains and disciplines. Given the interdisciplinary nature of ethical dataset curation, we seek to invite participants across a breadth of areas, including, but not limited to, HCI, AI, ML, STS, psychology, sociology, anthropology, law, policy, and ethics.

Before the workshop, we will invite participants to submit position papers (2–4 single spaced pages) via EasyChair in the ACM single-column format. Submissions will be reviewed by the program committee. In the position papers, participants will be asked to share their insights on the state of ethical large-scale dataset curation and thoughts on open areas for exploration. Submitted position papers will be assessed by the workshop organizers based not only on a paper's relevance to the workshop, but also on its ability to provoke discussion. To promote diversity of attendance, we will also include an alternative application option by using Google forms. With author permission, accepted position papers will be made available on the workshop website as workshop proceedings. Potential themes include, but are not limited to:

- Ethical principles and/or fairness definitions for AI datasets
- Challenges to collecting ethical datasets
- Dataset labor practices, including collection and annotation
- Identification and mitigation of dataset biases
- Best practices for dataset collection, maintenance, distribution, management, and use
- Implementations of practices and tools for ethical dataset development
- Data subject perspectives on data use, including from specific identity groups
- Legal practices and policy frameworks relevant to datasets
- Environmental or sustainability concerns surrounding datasets
- Tools and artifacts focused on improving responsibility of datasets

10 ORGANIZERS

The organizing committee comprises researchers, practitioners, and lawyers with backgrounds in HCI, ML, CV, NLP, algorithmic fairness, and social science with representation from both industry and academia. With a range of ethnic, cultural, and gender backgrounds, the committee brings extensive experience in dataset design, model training, and ethical guideline development.

Morgan Klaus Scheuerman is a Research Scientist at Sony AI within its AI Ethics team and a visiting scholar in Information Science at University of Colorado Boulder. He is interested in how identity characteristics are embedded into algorithmic infrastructures, like datasets, and how those permutations influence the entire system.

Dora Zhao is a PhD student at Stanford University. Her research focuses on uncovering, evaluating, and mitigating social biases in AI systems, primarily considering the computer vision and image-text domains. This includes improving dataset curation practices.

Jerone T. A. Andrews is a Research Scientist at Sony AI within its AI Ethics flagship project. His current research centers on human-centric computer vision, in particular, responsible data curation, bias detection and mitigation, and human-centric representation learning.

Abeba Birhane is a Senior Advisor in AI Accountability at Mozilla Foundation and an Adjunct Assistant Professor at Trinity College Dublin. She also serves on the UN Secretary-General's AI Advisory Body and the AI Advisory Council in Ireland. Her research focuses on AI accountability, with a particular focus on audits of AI models and training datasets.

Q. Vera Liao is an Associate Professor of Computer Science and Engineering at the University of Michigan. Her research interests are in human-AI interaction and responsible AI, with an overarching goal of bridging emerging AI technologies and human-centered design practices.

Georgia Panagiotidou is an Assistant Professor at King's College London. Her research examines how people interact with data and specifically how they handle issues such as biases, uncertainties and frictions when using data visualizations. She is increasingly interested in topics surrounding sustainability and climate justice as they relate to AI.

Pooja Chitre is a PhD student at Arizona State University. Her research is at the intersections of health policy, critical data studies, and postcolonial studies. Her present work focuses on understanding the data work required to make data actionable for healthcare organizations.

Kathleen Pine is an Associate Professor in the College of Health Solutions at Arizona State University. Her research examines design, implementation, and use of health information technologies, and has researched how increased demands for data work are reconfiguring healthcare practice.

Shawn Walker is an Assistant Professor in the School of Social and Behavioral Sciences at Arizona State University. His research focuses on mis and disinformation, and the challenges of collecting, analyzing, and preserving data from social media platforms. He also examines the use and implications of social media and web archives to train ML models.

Jieyu Zhao is an Assistant Professor at University of Southern California. Her research lies in detecting and mitigating societal biases in NLP and ML. Her current research focuses on examining, understanding, and reducing biases in large language models and promoting human-AI collaboration.

Alice Xiang is the Global Head of AI Ethics for Sony Group Corporation and Lead Research Scientist for Sony AI. Her current research focuses on best practices for ethical data curation and developing benchmarks and tools for human-centric AI technologies, particularly around addressing issues of diversity, transparency, and mitigating biases.

REFERENCES

[1] 2023. EU AI Act: First Regulation on Artificial Intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[2] Sanna J. Ali, Angèle Christin, Andrew Smart, and Riitta Katila. 2023. Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 217–226. https://doi.org/10.1145/3593013.3593990

[3] Antonio Aloisi. 2015. Commoditized Workers: Case Study Research on Labor Law Issues Arising from a Set of on-Demand/Gig Economy Platforms. *Comparative Labor Law & Policy Journal* 37 (2015).

[4] Jerone Andrews, Dora Zhao, William Thong, Apostolos Modas, Orestis Papakyriakopoulos, and Alice Xiang. 2023. Ethical Considerations for Responsible Data Curation. In *Thirty-Seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

[5] Jerone TA Andrews, Dora Zhao, William Thong, Apostolos Modas, Orestis Papakyriakopoulos, Shruti Nagpal, and Alice Xiang. 2023. Ethical considerations for collecting human-centric image datasets. *arXiv preprint arXiv:2302.03629* (2023).

[6] P Barger, TS Behrend, DJ Sharek, and EF Sinar. 2011. IO and the crowd: Frequently asked questions about using Mechanical Turk for research. *The Industrial-Organizational Psychologist* 49, 2 (2011), 11–17.

[7] Amirsiavosh Bashardoust, Stefan Feuerriegel, and Yash Raj Shrestha. 2024. Comparing the Willingness to Share for Human-generated vs. AI-generated Fake News. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2 (Nov. 2024), 489:1–489:21. https://doi.org/10.1145/3687028

[8] Jo Bates, Elli Gerakopoulou, and Alessandro Checco. 2023. Addressing Labour Exploitation in the Data Science Pipeline: Views of Precarious US-based Crowdworkers on Adversarial and Co-Operative Interventions. *Journal of Information, Communication and Ethics in Society* 21, 3 (Jan. 2023), 342–357. https://doi.org/10.1108/JICES-08-2022-0069

[9] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[10] Abeba Birhane, Sepehr Dehdashtian, Vinay Prabhu, and Vishnu Boddeti. 2024. The Dark Side of Dataset Scaling: Evaluating Racial Classification in Multimodal Models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1229–1244. https://doi.org/10.1145/3630106.3658968

[11] Abeba Birhane and Vinay Uday Prabhu. 2021. Large image datasets: A pyrrhic win for computer vision?. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1536–1546.

[12] Abeba Birhane and Vinay Uday Prabhu. 2021. Large Image Datasets: A Pyrrhic Win for Computer Vision?. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1536–1546. https://doi.org/10.1109/WACV48630.2021.00158

[13] Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe. 2021. Multimodal datasets: misogyny, pornography, and malignant stereotypes. *arXiv preprint arXiv:2110.01963* (2021).

[14] Alceu Bissoto, Catarina Barata, Eduardo Valle, and Sandra Avila. 2024. Even Small Correlation and Diversity Shifts Pose Dataset-Bias Issues. *Pattern Recognition Letters* 179 (March 2024), 87–93. https://doi.org/10.1016/j.patrec.2024.01.026

[15] Hannah Bleher and Matthias Braun. 2023. Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice. *Science and Engineering Ethics* 29, 3 (May 2023), 21. https://doi.org/10.1007/s11948-023-00443-3

[16] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings. In *Advances in Neural Information Processing Systems*, Vol. 29. Curran Associates, Inc.

[17] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, S Buch, Dallas Card, Rodrigo Castellon, Niladri S Chatterji, Annie S Chen, Kathleen A Creel, Jared Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren E Gillespie, Karan Goel, Noah D Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas F Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, O Khattab, Pang Wei Koh, Mark S Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D Manning, Suvir P Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Benjamin Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, J F Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Robert Reich, Hongyu Ren, Frieda Rong, Yusuf H Roohani, Camilo Ruiz, Jack Ryan, Christopher R'e, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishna Parasuram Srinivasan, Alex Tamkin, Rohan Taori, Armin W Thomas, Florian Tramèr, Rose E Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei A Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the Opportunities and Risks of Foundation Models. *ArXiv* abs/2108.0 (2021).

[18] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.

[19] Marco Cascella, Jonathan Montomoli, Valentina Bellini, and Elena Bignami. 2023. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *Journal of Medical Systems* 47, 1 (March 2023), 33. https://doi.org/10.1007/s10916-023-01925-4

[20] Boyang Chen, Zongxiao Wu, and Ruoran Zhao. 2023. From Fiction to Fact: The Growing Role of Generative AI in Business and Finance. *Journal of Chinese Economic and Business Studies* 21, 4 (Oct. 2023), 471–496. https://doi.org/10.1080/14765284.2023.2245279

[21] Youjin Choi, JaeYoung Moon, Kyung-Joong Kim, and Jin-Hyuk Hong. 2024. Exploring the Potential of Generative AI in Song-Signing. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '24)*. Association for Computing Machinery, New York, NY, USA, 816–820. https://doi.org/10.1145/3675094.3678378

[22] Jan Clusmann, Fiona R. Kolbinger, Hannah Sophie Muti, Zunamys I. Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P. Veldhuizen, Sophia J. Wagner, and Jakob Nikolas Kather. 2023. The Future Landscape of Large Language Models in Medicine. *Communications Medicine* 3, 1 (Oct. 2023), 1–8. https://doi.org/10.1038/s43856-023-00370-1

[23] Nicholas Kluge Corrêa, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, Rodrigo Mambrini, Luiza Galvão, Edmund Terem, and Nythamar de Oliveira. 2023. Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance. *Patterns* 4, 10 (Oct. 2023), 100857. https://doi.org/10.1016/j.patter.2023.100857

[24] Kate Crawford and Trevor Paglen. 2021. Excavating AI: The politics of images in machine learning training sets. *Ai & Society* 36, 4 (2021), 1105–1116.

[25] Nicolas Croce and Moh Musa. 2019. The new assembly lines: Why ai needs low-skilled workers too. https://www.weforum.org/agenda/2019/08/ai-low-skilled-workers/

[26] Roxana Daneshjou, Mary P. Smith, Mary D. Sun, Veronica Rotemberg, and James Zou. 2021. Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. *JAMA Dermatology* 157, 11 (Nov. 2021), 1362–1369. https://doi.org/10.1001/jamadermatol.2021.3129

[27] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[28] Devansh. 2023. Data Laundering: How Stability AI Managed to Get Millions of Copyrighted Artworks without Paying. . . .

[29] Mark Díaz, Ian Kivlichan, Rachel Rosen, Dylan Baker, Razvan Amironesei, Vinodkumar Prabhakaran, and Emily Denton. 2022. CrowdWorkSheets: Accounting for Individual and Collective Identities Underlying Crowdsourced Dataset Annotation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 2342–2351. https://doi.org/10.1145/3531146.3534647

[30] Vittoria Elliott. 2024. AI Tools Are Secretly Training on Real Images of Children. *Wired* (June 2024).

[31] Alessandro Fabris, Stefano Messina, Gianmaria Silvello, and Gian Antonio Susto. 2022. Tackling documentation debt: a survey on algorithmic fairness datasets. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–13.

[32] Michael Feffer, Nikolas Martelaro, and Hoda Heidari. 2023. The AI Incident Database as an Educational Tool to Raise Awareness of AI Harms: A Classroom Exploration of Efficacy, Limitations, & Future Improvements. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3617694.3623223

[33] Gabriele Ferri and Inte Gloerich. 2023. Risk and Harm: Unpacking Ideologies in the AI Discourse. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3571884.3603751

[34] Jade S. Franklin, Karan Bhanot, Mohamed Ghalwash, Kristin P. Bennett, Jamie McCusker, and Deborah L. McGuinness. 2022. An Ontology for Fairness Metrics. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*. Association for Computing Machinery, New York, NY, USA, 265–275. https://doi.org/10.1145/3514094.3534137

[35] Noa Garcia, Yusuke Hirota, Yankun Wu, and Yuta Nakashima. 2023. Uncurated image-text datasets: Shedding light on demographic bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6957–6966.

[36] Mitchell L Gordon, Michelle S Lam, Joon Sung Park, Kayur Patel, Jeff Hancock, Tatsunori Hashimoto, and Michael S Bernstein. 2022. Jury learning: Integrating dissenting voices into machine learning models. In *Conference on Human Factors in Computing Systems (CHI)*. 1–19.

[37] Mary L. Gray and Suri Siddharth. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.*

[38] Ben Green. 2021. The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice. *Journal of Social Computing* 2, 3 (Sept. 2021), 209–225. https://doi.org/10.23919/JSC.2021.0018

[39] Philipp Hacker. 2024. Sustainable AI Regulation. *Common Market Law Review* 61, 2 (April 2024).

[40] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[41] Kenji Hata, Ranjay Krishna, Li Fei-Fei, and Michael S Bernstein. 2017. A glimpse far into the future: Understanding long-term crowd worker quality. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 889–901.

[42] Yusuke Hirota, Yuta Nakashima, and Noa Garcia. 2022. Gender and racial bias in visual question answering datasets. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1280–1292.

[43] Yusuke Hirota, Yuta Nakashima, and Noa Garcia. 2022. Gender and Racial Bias in Visual Question Answering Datasets. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1280–1292. https://doi.org/10.1145/3531146.3533184 arXiv:2205.08148 [cs]

[44] Kenneth Holstein, Hal Daumé III, Miroslav Dudík, Hanna Wallach, and Jennifer Wortman Vaughan. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 16. https://doi.org/10.1145/3290605.3300830 arXiv:1812.05239v1

[45] Wiebke Hutiri, Orestis Papakyriakopoulos, and Alice Xiang. 2024. Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 359–376. https://doi.org/10.1145/3630106.3658911

[46] Oana Inel, Tim Draws, and Lora Aroyo. 2023. Collect, Measure, Repeat: Reliability Factors for Responsible AI Data Collection. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 11, 1 (Nov. 2023), 51–64. https://doi.org/10.1609/hcomp.v11i1.27547

[47] Lilly Irani. 2015. The cultural work of microwork. *New media & society* 17, 5 (2015), 720–739.

[48] Nitisha Jain, Mubashara Akhtar, Joan Giner-Miguelez, Rajat Shinde, Joaquin Vanschoren, Steffen Vogler, Sujata Goswami, Yuhan Rao, Tim Santos, Luis Oala, Michalis Karamousadakis, Manil Maskey, Pierre Marcenac, Costanza Conforti, Michael Kuchnik, Lora Aroyo, Omar Benjelloun, and Elena Simperl. 2024. A Standardized Machine-readable Dataset Documentation Format for Responsible AI. https://doi.org/10.48550/arXiv.2407.16883 arXiv:2407.16883 [cs]

[49] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 310–323. https://doi.org/10.1145/3531146.3533097

[50] Chenyan Jia, Michelle S. Lam, Minh Chau Mai, Jeffrey T. Hancock, and Michael S. Bernstein. 2024. Embedding Democratic Values into Social Media AIs via Societal Objective Functions. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 163:1–163:36. https://doi.org/10.1145/3641002

[51] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1, 9 (Sept. 2019), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[52] Yotam Kaplan and Ayelet Gordon-Tapiero. 2024. Generative AI Training as Unjust Enrichment.

[53] Chinmay Kulkarni, Tongshuang Wu, Kenneth Holstein, Q. Vera Liao, Min Kyung Lee, Mina Lee, and Hariharan Subramonyam. 2023. LLMs and the Infrastructure of CSCW. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 408–410. https://doi.org/10.1145/3584931.3608438

[54] Michelle S Lam, Mitchell L Gordon, Danaë Metaxa, Jeffrey T Hancock, James A Landay, and Michael S Bernstein. 2022. End-user audits: A system empowering communities to lead large-scale investigations of harmful algorithmic behavior. *proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34.

[55] Tina B. Lassiter and Kenneth R. Fleischmann. 2024. "Something Fast and Cheap" or "A Core Element of Building Trust"? - AI Auditing Professionals' Perspectives on Trust in AI. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2 (Nov. 2024), 424:1–424:22. https://doi.org/10.1145/3686963

[56] David Kuo Chuen Lee, Chong Guan, Yinghui Yu, and Qinxu Ding. 2024. A Comprehensive Review of Generative AI in Finance. *FinTech* 3, 3 (Sept. 2024), 460–478. https://doi.org/10.3390/fintech3030025

[57] Shayne Longpre, Robert Mahari, Ariel Lee, Campbell Lund, Hamidah Oderinwale, William Brannon, Nayan Saxena, Naana Obeng-Marnu, Tobin South, Cole Hunter, Christopher Klamm, Hailey Schoelkopf, Nikhil Singh, Manuel Cherep, Mustafa Anis, An Dinh, Caroline Chitongo, Da Yin, Damien Sileo, Deividas Mataciunas, Diganta Misra, Emad Alghamdi, Enrico Shippole, Jianguo Zhang, Joanna Materzynska, Kun Qian, Kush Tiwary, Lester Miranda, Manan Dey, Minnie Liang, Niklas Muennighoff, Seonghyeon Ye, Seungone Kim, Shrestha Mohanty, Vivek Sharma, Vu Minh Chien, Xuhui Zhou, Yizhi Li, Caiming Xiong, Luis Villa, Stella Biderman, Hanlin Li, Daphne Ippolito, Sara Hooker, and Jad Kabbara. 2024. Consent in Crisis: The Rapid Decline of the AI Data Commons. (2024).

[58] Alexandra Luccioni and Joseph Viviano. 2021. What's in the box? an analysis of undesirable content in the Common Crawl corpus. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 182–189.

[59] Jabez Magomere, Shu Ishida, Tejumade Afonja, Aya Salama, Daniel Kochin, Foutse Yuehgoh, Imane Hamzaoui, Raesetje Sefala, Aisha Alaagib, Elizaveta Semenova, Lauren Crais, and Siobhan Mackenzie Hall. 2024. You Are What You Eat? Feeding Foundation Models a Regionally Diverse Food Dataset of World Wide Dishes. https://doi.org/10.48550/arXiv.2406.09496 arXiv:2406.09496 [cs]

[60] Nicolas Malevé. 2020. On the data set's ruins. *AI & SOCIETY* (2020), 1–15.

[61] Kate McCandless. 2021. Just Because the Data Is There, It Doesn't Mean It's Yours to Take : Exploring User, Researcher and Review Board Perceptions in Twitter Data Research. *Emerging Library & Information Perspectives* 4, 1 (July 2021), 34–61. https://doi.org/10.5206/elip.v4i1.13554

[62] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application* 8, 1 (March 2021), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

[63] Luke Munn. 2023. The Uselessness of AI Ethics. *AI and Ethics* 3, 3 (Aug. 2023), 869–877. https://doi.org/10.1007/s43681-022-00209-w

[64] Keziah Naggita, Julienne LaChance, and Alice Xiang. 2023. Flickr Africa: Examining Geo-Diversity in Large-Scale, Human-Centric Visual Data. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. Association for Computing Machinery, New York, NY, USA, 520–530. https://doi.org/10.1145/3600211.3604659

[65] Nadia Nahar, Jenny Rowlett, Matthew Bray, Zahra Abba Omar, Xenophon Papademetris, Alka Menon, and Christian Kästner. 2024. Regulating Explainability in Machine Learning Applications – Observations from a Policy Design Experiment. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 2101–2112. https://doi.org/10.1145/3630106.3659028

[66] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. 2023. Co-Design of Human-centered, Explainable AI for Clinical Decision Support. *ACM Trans. Interact. Intell. Syst.* 13, 4 (Dec. 2023), 21:1–21:35. https://doi.org/10.1145/3587271

[67] Orestis Papakyriakopoulos, Anna Seo Gyeong Choi, William Thong, Dora Zhao, Jerone Andrews, Rebecca Bourke, Alice Xiang, and Allison Koenecke. 2023. Augmented Datasheets for Speech Datasets and Ethical Decision-Making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 881–904. https://doi.org/10.1145/3593013.3594049

[68] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (2021).

[69] Addie Payne Morgan, Bryan F. Howell, and Grace Kilbourn-Barber. 2024. CO-DESIGN AND ARTIFICIAL INTELLIGENCE: A METHOD TO EMPOWER END-USERS IN VISUAL COMMUNICATION. In *DS 131: Proceedings of the International Conference on Engineering and Product Design Education (E&PDE 2024)*. 557–562. https://doi.org/10.35199/EPDE.2024.94

[70] Kenny Peng, Arunesh Mathur, and Arvind Narayanan. 2021. Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers. (Aug. 2021). arXiv:2108.02922

[71] Kenneth L Peng, Arunesh Mathur, and Arvind Narayanan. 2021. Mitigating dataset harms requires stewardship: Lessons from 1000 papers. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

[72] Amit Pinchevski. 2023. Social media's canaries: content moderators between digital labor and mediated trauma. *Media, Culture & Society* 45, 1 (2023), 212–221.

[73] Anaïs Rességuier and Rowena Rodrigues. 2020. AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics. *Big Data & Society* 7, 2 (July 2020), 2053951720942541. https://doi.org/10.1177/2053951720942541

[74] Kat Roemmich, Shanley Corvite, Cassidy Pyle, Nadia Karizat, and Nazanin Andalibi. 2024. Emotion AI Use in U.S. Mental Healthcare: Potentially Unjust and Techno-Solutionist. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 47:1–47:46. https://doi.org/10.1145/3637324

[75] Niamh Rowe. 2023. 'It's destroyed me completely': Kenyan moderators decry toll of training of AI models. *The Guardian* (2023). https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

[76] Pamela Samuelson. 2023. Generative AI Meets Copyright. *Science* 381, 6654 (July 2023), 158–161. https://doi.org/10.1126/science.adi0656

[77] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[78] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade W. Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa R. Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. 2022. LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models. In *Thirty-Sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

[79] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Foregrounding Data in Co-Design – An Exploration of How Data May Become an Object of Design. *International Journal of Human-Computer Studies* 143 (Nov. 2020), 102505. https://doi.org/10.1016/j.ijhcs.2020.102505

[80] Boaz Shmueli, Jan Fell, Soumya Ray, and Lun-Wei Ku. 2021. Beyond Fair Pay: Ethical Implications of NLP Crowdsourcing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 3758–3769. https://doi.org/10.18653/v1/2021.naacl-main.295 arXiv:2104.10097 [cs]

[81] Luke Stark. 2023. Breaking Up (with) AI Ethics. *American Literature* 95, 2 (June 2023), 365–379. https://doi.org/10.1215/00029831-10575148

[82] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 3645–3650. https://doi.org/10.18653/v1/P19-1355

[83] David Thiel. 2023. Investigation Finds AI Image Generation Models Trained on Child Abuse. https://fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse.

[84] Eddie L. Ungless, Björn Ross, and Anne Lauscher. 2023. Stereotypes and Smut: The (Mis)Representation of Non-cisgender Identities by Text-to-Image Models. (May 2023). arXiv:2305.17072

[85] Priyan Vaithilingam, Ian Arawjo, and Elena L. Glassman. 2024. Imagining a Future of Designing with AI: Dynamic Grounding, Constructive Negotiation, and Sustainable Motivation. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 289–300. https://doi.org/10.1145/3643834.3661525

[86] Hamish van der Ven, Diego Corry, Rawie Elnur, Viola Jasmine Provost, and Muh Syukron. 2024. Generative AI and Social Media May Exacerbate the Climate Crisis. *Global Environmental Politics* 24, 2 (May 2024), 9–18. https://doi.org/10.1162/glep_a_00747

[87] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks Posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 214–229. https://doi.org/10.1145/3531146.3533088

[88] Lauren Wilcox, Robin Brewer, and Fernando Diaz. 2023. AI Consent Futures: A Case Study on Voice Data Collection with Clinicians. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 316:1–316:30. https://doi.org/10.1145/3610107

[89] Adrienne Williams, Milagros Miceli, and Timnit Gebru. 2022. The Exploited Labor behind Artificial Intelligence. *NOEMA* (2022), 1–11.

[90] Wenying Wu, Pavlos Protopapas, Zheng Yang, and Panagiotis Michalatos. 2020. Gender Classification and Bias Mitigation in Facial Images. In *Proceedings of the 12th ACM Conference on Web Science (WebSci '20)*. Association for Computing Machinery, New York, NY, USA, 106–114. https://doi.org/10.1145/3394231.3397900

[91] Chloe Xiang. 2023. OpenAI and Microsoft Sued for $3 Billion Over Alleged ChatGPT 'Privacy Violations'.

[92] Meg Young, Luke Rodriguez, Emily Keller, Feiyang Sun, Boyang Sa, Jan Whittington, and Bill Howe. 2019. Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 191–200. https://doi.org/10.1145/3287560.3287577

[93] Brian Hu Zhang, Blake Lemoine, and Margaret Mitchell. 2018. Mitigating Unwanted Biases with Adversarial Learning. In *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM Press, New York, New York, USA, 335–340. https://doi.org/10.1145/3278721.3278779 arXiv:1801.07593

[94] Dora Zhao, Jerone Andrews, Orestis Papakyriakopoulos, and Alice Xiang. 2024. Position: Measure Dataset Diversity, Don't Just Claim It. In *Proceedings of the 41st International Conference on Machine Learning*. PMLR, 60644–60673.

[95] Dora Zhao, Morgan Klaus Scheuerman, Pooja Chitre, Jerone T. A. Andrews, Georgia Panagiotidou, Shawn Walker, Kathleen H. Pine, and Alice Xiang. 2024. A Taxonomy of Challenges to Curating Fair Datasets. In *NeurIPS 2024*. https://doi.org/10.48550/arXiv.2406.06407 arXiv:2406.06407 [cs]

[96] Dora Zhao, Angelina Wang, and Olga Russakovsky. 2021. Understanding and evaluating racial biases in image captioning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14830–14840.

[97] Michael Zimmer. 2010. "But the Data Is Already Public": On the Ethics of Research in Facebook. *Ethics and Information Technology* 12, 4 (Dec. 2010), 313–325. https://doi.org/10.1007/s10676-010-9227-5

[98] Matthew Zook, Solon Barocas, Danah Boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A Koenig, Jacob Metcalf, et al. 2017. Ten simple rules for responsible big data research. , e1005399 pages.