Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research

Other titles in Foundations and Trends® in Computer Graphics and Vision

Computational Imaging Through Atmospheric Turbulence
Stanley H. Chan and Nicholas Chimitt
ISBN: 978-1-63828-999-9

Towards Better User Studies in Computer Graphics and Vision
Zoya Bylinskii, Laura Herman, Aaron Hertzmann, Stefanie Hutka and Yile Zhang
ISBN: 978-1-63828-172-6

An Introduction to Neural Data Compression
Yibo Yang, Stephan Mandt and Lucas Theis
ISBN: 978-1-63828-174-0

Learning-based Visual Compression
Ruolei Ji and Lina J. Karam
ISBN: 978-1-63828-112-2

Vision-Language Pre-Training: Basics, Recent Advances, and Future Trends
Zhe Gan, Linjie Li, Chunyuan Li, Lijuan Wang, Zicheng Liu and Jianfeng Gao
ISBN: 978-1-63828-132-0

Semantic Image Segmentation: Two Decades of Research
Gabriela Csurka, Riccardo Volpi and Boris Chidlovskii
ISBN: 978-1-63828-076-7

Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research

Timnit Gebru
The Distributed AI Research Institute (DAIR)
timnit@dair-institute.org

Remi Denton
Google Research
dentone@google.com

the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

T. Gebru and R. Denton. *Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research.* Foundations and Trends® in Computer Graphics and Vision, vol. 16, no. 3, pp. 215–321, 2024.

ISBN: 978-1-63828-355-3

Foundations and Trends® in Computer Graphics and Vision
Volume 16, Issue 3, 2024

Editorial Board

Editorial Scope

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2024, Volume 16, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research

Timnit Gebru[1] and Remi Denton[2]

[1] *The Distributed AI Research Institute (DAIR), USA; timnit@dair-institute.org*
[2] *Google Research, USA; dentone@google.com*

ABSTRACT

The field of computer vision is now a multi-billion dollar enterprise, with its use in surveillance applications driving this large market share. In the last six years, computer vision researchers have started to discuss the risks and harms of some of these systems, mostly using the lens of fairness introduced in the machine learning literature to perform this analysis. While this lens is useful to uncover and mitigate a narrow segment of the harms that can be enacted through computer vision systems, it is only one of the toolkits that researchers have available to uncover and mitigate the harms of the systems they build.

In this monograph, we discuss a wide range of risks and harms that can be enacted through the development and deployment of computer vision systems. We also discuss some

Timnit Gebru and Remi Denton (2024), "Beyond Fairness in Computer Vision: A Holistic Approach to Mitigating Harms and Fostering Community-Rooted Computer Vision Research", Foundations and Trends® in Computer Graphics and Vision: Vol. 16, No. 3, pp 215–321. DOI: 10.1561/0600000102.

existing technical approaches to mitigating these harms, as well as the shortcomings of these mitigation strategies. Then, we introduce computer vision researchers to harm mitigation strategies proposed by journalists, human rights activists, individuals harmed by computer vision systems, and researchers in disciplines ranging from sociology to physics. We conclude the monograph by listing principles that researchers can follow to build what we call community-rooted computer vision tools in the public interest, and give examples of such research directions. We hope that this monograph can serve as a starting point for researchers exploring the harms of current computer vision systems and attempting to steer the field into community-rooted work.

1

Introduction

Computer vision is no longer a purely academic endeavor, instead it is a field with significant industry market share. According to Fortune Business Insights, the global image recognition market size is projected to reach USD $178.13 billion by 2032,[1] with the "active adoption of facial recognition technologies to screen people amid the current COVID-19 pandemic ensuring "stable growth" of the market."[2] This market share estimate does not include the data annotation industry which is projected to reach $13.2 billion in market value by 2030.[3] Beyond industry applications, a number of computer vision researchers have also created tools in the public interest. For example, Sefala *et al.* (2021) analyzed satellite imagery to study the impacts of discriminatory

[1] https://www.fortunebusinessinsights.com/industry-reports/image-recognition-market-101855.

[2] https://www.globenewswire.com/news-release/2021/12/21/2355754/0/en/Image-Recognition-Market-to-Hit-86-32-Billion-by-2027-Rapid-Advancements-in-Face-Detection-Technologies-to-Drive-Market-Growth-Fortune-Business-Insights.html.

[3] https://www.businesswire.com/news/home/20220824005423/en/The-Worldwide-Data-Annotation-Tools-Industry-is-Expected-to-Reach-13.2-Billion-by-2030—ResearchAndMarkets.com.

policies, and Mutembesa *et al.* (2019) created a mobile app to help farmers monitor plant health on their farms.

In spite of some efforts to create computer vision tools in the public interest, the field's popularity and increase in industry and government use has led to negative consequences to marginalized groups, some of which have been discussed within the computer vision community, but many of which have not. This monograph gives an overview of the reported harms of computer vision systems and works that resist, refuse, mitigate or slow down these harms and provide an alternate path. We take a broad approach to surveying works, covering not only scholarship within computer vision and adjacent fields, but the important works of scholars in fields such as sociology, science and technology studies, communications, history, and surveillance studies who have done foundational work to contextualize the societal impacts of computer vision systems. Journalists, civil rights organizations and individuals who have been negatively impacted by computer vision systems have also played a major role in uncovering some of its harms, and workers have organized to curb a number of harmful use cases.

The rest of the monograph is structured as follows. Section 2 is a short positionality statement, following one of our recommendations to computer vision researchers (see Section 4.3). Section 3 gives an overview of the negative impacts that computer vision technologies can have on individuals, communities, and society. Our overview draws terminology and typologies from recent scholarship that has offered taxonomies of harm relating to machine learning technologies (e.g., Katzman *et al.*, 2023; Wang *et al.*, 2022a; Shelby *et al.*, 2023), but focuses specifically on harms associated with computer vision. We specifically use the taxonomy proposed by Shelby *et al.* (2023), which categorizes the harms posed by automated systems into 5 categories: representational harms, allocative harms, quality of service harms, interpersonal harms and social system harms. These taxonomies are not mutually exclusive however, and many examples can fall into multiple categories. Alongside our summary of reported harms, we discuss efforts to uncover and mitigate these harms, as well as the shortcomings of harm reduction strategies that solely depend on technical interventions.

We then offer forward-looking guidance and recommendations for computer vision researchers and practitioners that we hope will guide the field towards more equitable and just outcomes. Section 4 summarizes actions individual computer vision researchers and practitioners can take, such as following recommended responsible AI frameworks, documentation standards, and other principles. Section 5 discusses the properties of computer vision systems that result in tools that help marginalized communities rather than surveil them, and research practices that enable the development of these types of tools. Section 6 takes a bigger-picture look at the institutions and power structures that scaffold computer vision research and development practices, and discusses avenues that researchers and practitioners can take to institute systemic change to encourage what we call community-rooted computer vision research.

This monograph is predominantly written for computer vision students, researchers, and practitioners. Within the computer vision community, and the machine learning community more generally, harms arising from computer vision and/or machine learning systems are often discussed using the narrow lenses of model fairness and bias. While this monograph does not specifically focus on these concepts, we give some examples of interventions that rely on them, and also elaborate on why these types of technical interventions can be limited. We encourage our readers to familiarize themselves with techniques used by the ML fairness community by reading the comprehensive book *Fairness and Machine Learning* (Barocas *et al.*, 2023), reviewing the many measures of fairness that have been introduced over the years (Narayanan, 2018), and reading surveys of debiasing methods used in neural-network based vision and language research (e.g., Parraga *et al.*, 2023).

In this monograph, we intentionally survey work that may be less familiar to computing researchers, including scholarship from a broad set of non-computing disciplines, and reports from journalists, human rights experts, and individuals directly impacted by computer vision systems. In doing so, we aim to introduce computer vision researchers to the broader discourse on the ethical implications of computer vision research, and deepen their awareness of the harms that can stem from the research, development, and deployment of a number of computer

vision technologies. We discuss interventions at the data and model level that can mitigate some of these harms (e.g., removing hate speech from an image captioning dataset to mitigate the risk of a model captioning an image with hateful language). However, many of the harms we discuss are ones that do not have purely "technical" interventions, because we aim to equip computer vision researchers and practitioners with a broader set of tools to understand and mitigate harms arising from the systems they build–including having the option of not building certain systems.

We recognize that our readers may feel that they have limited agency to prevent the negative impacts detailed in this monograph, particularly the societal-level harms discussed in Section 3.4 that may be understood as being "further" from model development and deployment practices. But we hope that the many examples of interventions provided in this monograph from critical interdisciplinary perspectives can serve as inspiration to students, researchers and practitioners in computer vision to move the field away from the surveillance of marginalized communities, and into serving their needs instead. We also recognize that due to the interdisciplinary nature of this monograph, many readers may be unfamiliar with some terminology and concepts that we introduce. As a result, we have striven to define terms external to the field of computer vision, and provide references that can be used as a foundation for our readers' future learning.

References

Agre, P. E. (1997). "Toward a critical technical practice: Lessons learned in trying to reform AI". In: *Social Science, Technical Systems, and Cooperative Work*. Ed. by G. Bowker, G. Bowker, S. L. Star, L. Gasser, and W. Turner. Psychology Press.

Agüera y Arcas, B., M. Mitchell, and A. Todorov (2017). "Physiognomy's new clothes". URL: https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a.

Aizeki, M., G. Boyce, T. Miller, J. Nevins, and M. Ticktin (2021). "Smart borders or a humane world?" The Immigrant Defense Project's Surveillance, Tech & Immigration Policing Project, and the Transnational Institute.

Aizeki, M., M. Mahmoudi, and C. Schupfer eds. (2024). *Resisting Borders and Technologies of Violence*. New York, NY: Haymarket.

Ajunwa, I. (2021). "Automated video interviewing as the new phrenology". *Berkeley Technology Law Journal*. 36: 1173.

Alkhatib, A. (2019). "Anthropological/artificial intelligence and the HAI". URL: https://ali-alkhatib.com/blog/anthropological-intelligence.

Alvi, M., A. Zisserman, and C. Nellåker (2018). "Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Amnesty International (2023). "Israel and occupied palestinian territories: Automated apartheid: How facial recognition fragments, segregates and controls palestinians in the OPT. Amnesty International". URL: https://www.amnesty.org/en/documents/mde15/6701/2023/en/.

Anderson, S. (2020). "The alt-right manipulated my comic. Then A.I. claimed it". *New York Times*.

Andrus, M., E. Spitzer, J. Brown, and A. Xiang (2021). "What we can't measure, we can't understand: Challenges to demographic data procurement in the pursuit of fairness". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 249–260.

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

Asano, Y. M., C. Rupprecht, A. Zisserman, and A. Vedaldi (2021). "PASS: An ImageNet replacement for self-supervised pretraining without humans". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*.

Ashurst, C., E. Hine, P. Sedille, and A. Carlier (2022). "AI ethics statements: Analysis and lessons learnt from NeurIPS broader impact statements". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 2047–2056. DOI: 10.1145/3531146.3533780.

Ayyub, R. (2018). "I was the victim of a deepfake porn plot intended to silence me". *Huffington Post*. 18(2022). URL: https://www.huffpost.com/archive/in/entry/deepfake-porn_in_5c1201cfe4b0508b213746bd.

Balakrishnan, G., Y. Xiong, W. Xia, and P. Perona (2020). "Towards causal benchmarking of bias in face analysis algorithms". *Deep Learning-Based Face Analytics*: 327–359.

Barocas, S., K. Crawford, A. Shapiro, and H. Wallach (2017). "The problem with bias: Allocative versus representational harms in machine learning". In: *Proceedings of SIGCIS*. Philadelphia, PA.

Barocas, S., A. Guo, E. Kamar, J. Krones, M. R. Morris, J. W. Vaughan, W. D. Wadsworth, and H. Wallach (2021). "Designing disaggregated evaluations of AI systems: Choices, considerations, and tradeoffs". In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 368–378.

Barocas, S., M. Hardt, and A. Narayanan (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.

Barrett, L. F., R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak (2019). "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements". *Psychological Science in the Public Interest*. 20(1): 1–68.

Bellamy, R. K. E., K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, S. Nagar, K. N. Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang (2018). "AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias". *ArXiv, abs/1810.01943*.

Bender, E. M. (2023). "Cleaning up a baby peacock sullied by a non-information spill". In: *Medium*. URL: https://medium.com/@emilymenonbender/cleaning-up-a-baby-peacock-sullied-by-a-non-information-spill-d2e2aa642134.

Bender, E. M. and B. Friedman (2018). "Data statements for natural language processing: Toward mitigating system bias and enabling better science". *Transactions of the Association for Computational Linguistics*. 6: 587–604. DOI: 10.1162/tacl_a_00041.

Bender, E. M., T. Gebru, A. McMillan-Major, and S. 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 (FAccT '21)*. New York, NY, USA: Association for Computing Machinery. 610–623. DOI: 10.1145/3442188.3445922.

Bengio, S. and I. D. Raji (2021). "A retrospective on the NeurIPS 2021 ethics review process". *NeurIPS Blog*. URL: https://blog.neurips.cc/2021/12/03/a-retrospective-on-the-neurips-2021-ethics-review-process/.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley and Sons.

Benjamin, R. (2020). *Reimagining the Default Settings of Technology and Society*. ICLR Keynote.

Benjamin, R. (2024). *Imagination: A Manifesto (A Norton Short)*. WW Norton and Company.

Bennett, C. L., C. Gleason, M. K. Scheuerman, J. P. Bigham, A. Guo, and A. To (2021). "It's complicated: Negotiating accessibility and (mis) representation in image descriptions of race, gender, and disability". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

Bernstein, M. S., M. Levi, D. Magnus, and Q. Waeis (2021). "Ethics and society review: Ethics reflection as a precondition to research funding". In: *Proceedings of the National Academy of Science*.

Bianchi, F., P. Kalluri, E. Durmus, F. Ladhak, M. Cheng, D. Nozza, T. Hashimoto, D. Jurafsky, J. Zou, and A. Caliskan (2023). "Easily accessible text-to-image generation amplifies demographic stereotypes at large scale". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1493–1504.

Birhane, A., W. Isaac, V. Prabhakaran, M. Diaz, M. C. Elish, I. Gabriel, and S. Mohamed (2022a). "Power to the people? Opportunities and challenges for participatory AI". In: *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.

Birhane, A., P. Kalluri, D. Card, W. Agnew, R. Dotan, and M. Bao (2022b). "The values encoded in machine learning research". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 173–184. DOI: 10.1145/3531146.3533083.

Birhane, A. and V. U. Prabhu (2020). "Large image datasets: A pyrrhic win for computer vision?" In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 1536–1546.

Birhane, A., V. U. Prabhu, and E. Kahembwe (2021). "Multimodal datasets: Misogyny, pornography, and malignant stereotypes". *arXiv preprint arXiv:2110.01963*.

Black, E. (2001). *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation*. Dialog press.

Boyd, K. L. (2021). "Datasheets for datasets help ML engineers notice and understand ethical issues in training data". *Proceedings of the ACM on Human-Computer Interaction 5, CSCW2, Article 438 (October 2021)*: 27. DOI: 10.1145/3479582.

Brock, A. (2022). "It's not the data: Weak tie algorithmic sociality and digital culture". In: *Keynote talk, ACM Conference on Fairness, Accountability, and Transparency*. URL: https://www.youtube.com/watch?v=Nx3N1961t08.

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.

Buolamwini, J. and T. Gebru (2018). "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, Accountability and Transparency*. PMLR. 77–91.

Byler, D. (2021a). "The covid tech that is intimately tied to China's surveillance state". *MIT Technology Review*. URL: https://www.technologyreview.com/2021/10/11/1036582/darren-byler-xinjiang-china-uyghur-surveillance/.

Byler, D. (2021b). "In the camps: China's high-tech penal colony. Columbia global reports". *JSTOR*. URL: http://www.jstor.org/stable/j.ctv2dzzqqm.

Catanzariti, B., S. Chandhiramowuli, S. Mohamed, S. Natarajan, S. Prabhat, N. Raval, A. S. Taylor, and D. Wang (2021). "The global labours of AI and data intensive systems". In: *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA. 319–322. DOI: 10.1145/3462204.3481725.

Chowdhury, R. (2022). "The growing resistance to Biden's 'smart' border — and what it means for climate refugees". *Grist*. URL: https://grist.org/fix/advocacy/the-growing-resistance-to-bidens-smart-border-and-what-it-means-for-climate-refugees.

Cifor, M., P. Garcia, T. L. Cowan, J. Rault, T. Sutherland, A. Chan, J. Rode, A. L. Hoffmann, N. Salehi, and L. Nakamura (2019). "Feminist data manifest-no". URL: https://www.manifestno.com/.

Cioffi *et al.* (2021). *Chased Away and Left to Die: How a National Security Approach to Uganda's National ID Has Led to Wholesale Exclusion of Women and Older Persons*. URL: https://www.google.com/url?q=https://chrgj.org/wp-content/uploads/2021/06/CHRGJ-Report-Chased-Away-and-Left-to-Die.pdf&sa=D&source=editors&ust=1715560526721500&usg=AOvVaw0vgKYKGwx-wYpE6RpihgRL.

Coding Rights (2022). "Legislators from all regions of Brazil present bills to ban facial recognition in public spaces". *Medium*. URL: https://medium.com/codingrights/legislators-from-all-regions-of-brazil-present-bills-to-ban-facial-recognition-in-public-spaces-31d8da0d3822.

Collins, E. (2022). "Abolishing the evidence-based paradigm". *Brigham Young University Law Review*. 48(2): 403–462.

Collins, P. H. (1990). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge.

Cooper, N., T. Horne, G. R. Hayes, C. Heldreth, M. Lahav, J. Holbrook, and L. Wilcox (2022). "A systematic review and thematic analysis of community-collaborative approaches to computing research". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

Corbett, E., E. Denton, and S. Erete (2023). "Power and public participation in AI". In: *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–13.

Crawford, K. and T. Paglen (2021). "Excavating AI: The politics of images in machine learning training sets". *AI and Society*. 36(4): 1105–1116.

Creager, E., D. Madras, J.-H. Jacobsen, M. Weis, K. Swersky, T. Pitassi, and R. Zemel (2019). "Flexibly fair representation learning by disentanglement". In: *International Conference on Machine Learning*. PMLR. 1436–1445.

Crenshaw, K. (2016). "The urgency of intersectionality". *TED*.

Crenshaw, K. W. (1997). "Mapping the margins: Intersectionality, identity politics, and violence against women of color". In: *The Public Nature of Private Violence*. Routledge. 93–118.

Cummings, M. L. (2015). "Automation bias in intelligent time critical decision support systems". In: *Decision Making in Aviation*. Ed. by D. Harris.

Curtis, M. and T. Jones (2017). "Honest accounts 2017—How the world profits from Africa's wealth". *Curtis Research*. URL: https://curtisresearch.org/honest-accounts-2017-how-the-world-profits-from-africas-wealth/.

Cutler, M., E. Keith, M. Kleinman, M. Ky, Y. Navarrete, K. Pellar, D. Ramesh, and E. Welsby (2024). *Community Partnerships Playbook: How to Create Equitable Partnerships Between Technical and Community Experts*. Science, Technology, and Public Policy Program Gerald R. Ford School of Public Policy University of Michigan. URL: https://stpp.fordschool.umich.edu/research/research-report/community-partnerships-playbook.

Dai, J., J. Wu, B. Saghafi, J. Konrad, and P. Ishwar (2015). "Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 68–76.

Das, M., A. J. Fiannaca, M. R. Morris, S. K. Kane, and C. L. Bennett (2024). "From provenance to aberrations: Image creator and screen reader user perspectives on alt text for AI-generated images". In: *ACM CHI Conference on Human Factors in Computing Systems*.

De Vries, T., I. Misra, C. Wang, and L. Van der Maaten (2019). "Does object recognition work for everyone?" In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 52–59.

Del Valle, G. (2022). "The most surveilled place in America". *The Verge*. URL: https://www.theverge.com/c/23203881/border-patrol-wall-surveillance-tech.

Delgado, F., S. Yang, M. Madaio, and Q. Yang (2023). "The participatory turn in AI design: Theoretical foundations and the current state of practice". In: *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–23.

Demopoulos, A. (2022). "The inherent misogyny of AI portraits – Amelia Earhart rendered naked on a bed". *The Guardian*. URL: https://www.theguardian.com/us-news/2022/dec/09/lensa-ai-portraits-misogyny.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009). "ImageNet: A large-scale hierarchical image database". In: *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA. 248–255. DOI: 10.1109/CVPR.2009.5206848.

Denton, R., B. Hutchinson, M. Mitchell, T. Gebru, and A. Zaldivar (2019). "Image counterfactual sensitivity analysis for detecting unintended bias". In: *CVPR Workshop on Fairness Accountability Transparency and Ethics in Computer Vision*.

Dersso, S. A. (2022). "Brace yourselves for a new cold war". *Mail and Guardian*. URL: https://mg.co.za/africa/2022-02-04-brace-yourselves-for-a-new-cold-war/.

Diaz, M., S. Dev, E. Reif, E. Denton, and V. Prabhakaran (2023). "Developing a conceptual framework for analyzing people in unstructured data". In: *NeurIPS Workshop on Socially Responsible Language Modelling Research*.

Díaz, M., I. Kivlichan, R. Rosen, D. Baker, R. Amironesei, V. Prabhakaran, and E. Denton (2022). "CrowdWorkSheets: Accounting for individual and collective identities underlying crowdsourced dataset annotation". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 2342–2351. DOI: 10.1145/3531146.3534647.

Dobson, J. E. (2023). *The Birth of Computer Vision*. U of Minnesota Press.

Dodge, J., T. Prewitt, R. T. des Combes, E. Odmark, R. Schwartz, E. Strubell, A. S. Luccioni, N. A. Smith, N. DeCario, and W. Buchanan (2022). "Measuring the carbon intensity of AI in cloud instances". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 1877–1894. DOI: 10.1145/3531146.3533234.

Dwork, C. (2006a). "Differential privacy". In: *Automata, Languages and Programming*. Ed. by M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener.

Dwork, C. (2006b). "Differential privacy". In: *International Colloquium on Automata, Languages, and Programming*. Berlin, Heidelberg: Springer Berlin Heidelberg. 1–12.

Dwork, C., M. Hardt, T. Pitassi, O. Reingold, and R. Zemel (2012). "Fairness through awareness". In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. 214–226.

Dwork, C., N. Immorlica, A. T. Kalai, and M. Leiserson (2018). "Decoupled classifiers for group-fair and efficient machine learning". In: *Conference on Fairness, Accountability and Transparency*. PMLR. 119–133.

Feathers, T. (2020). "Car companies want to monitor your every move with emotion-detecting AI". *Vice*.

Fergusson, G., C. Fitzgerald, C. Frascella, M. Iorio, T. McBrien, C. Schroeder, B. Winters, and E. Zhou (2023). *Generating Harms: Generative AI's Impact and Paths Forward*. Electronic Privacy Information Center.

Franks, M. A. and A. E. Waldman (2019). "Sex, lies, and videotape: Deep fakes and free speech delusions". *Maryland Law Review*. 78.

Fraser, K. C., I. Nejadgholi, and S. Kiritchenko (2023). "A friendly face: Do text-to-image systems rely on stereotypes when the input is under-specified?" In: *The AAAI-23 Workshop on Creative AI Across Modalities*.

Friedman, D. and A. B. Dieng (2023). "The vendi score: A diversity evaluation metric for machine learning". *Transactions on Machine Learning Research*.

Frome, A., G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent (2009). "Large-scale privacy protection in google street view". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE. 2373–2380.

Fry, H. (2018). "What algorithms know about you based on your grocery cart and more strange tales from the wild west of data". *Medium*. URL: https://onezero.medium.com/what-algorithms-know-about-you-based-on-your-grocery-cart-b364e0ba545c.

Gangadharan, S. (2020). "Context, research, refusal: Perspectives on abstract problem-solving". In: *Invited talk, Towards Trustworthy ML: Rethinking Security and Privacy for ML, International Conference on Learning Representations*. URL: https://trustworthyiclr20.github.io/Gangadharan_ICLR_trustworthy_ML.pdf.

Garvie, C., A. Bedoya, and J. Frankle (2016). *The Perpetual Line-Up: Unregulated Police Face Recognition in America*. Georgetown Law, Center on Privacy and Technology. URL: https://www.perpetuallineup.org/.

Garvie, C. and L. M. Moy (2019). *America Under Watch: Face Surveillance in the United States*. Georgetown Law, Center on Privacy and Technology. URL: https://www.americaunderwatch.com/.

Gebru, T. (2021). *Timnit Gebru Keynote at UC Berkeley School of Information Commencement 2022*. URL: https://www.youtube.com/watch?v=VRsrcRDMT2w.

Gebru, T. (2022). "The quest for ethical artificial intelligence: A conversation with Timnit Gebru". In: *Harvard Radcliffe Institute Science Lecture Series on AI*. URL: https://www.youtube.com/watch?v=b_--xrN3eso.

Gebru, T. (2023). "Deep learning indaba talk". URL: https://www.youtube.com/watch?v=5cm_FvHmtVI&t=1s.

Gebru, T. (2024). "#TeamHuman: Community-rooted AI research". In: *Deep Learning Indaba*. URL: https://www.youtube.com/watch?v=5cm_FvHmtVI&t=1s.

Gebru, T. and R. Denton (2020a). "Computer vision in practice: Who is benefiting and who is being harmed? Tutorial on fairness accountability transparency and ethics in computer vision". URL: https://youtu.be/0sBE5OyD7fk.

Gebru, T. and R. Denton (2020b). "Tutorial on fairness accountability transparency and ethics in computer vision at CVPR". URL: https://sites.google.com/corp/view/fatecv-tutorial/.

Gebru, T., J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford (2021). "Datasheets for datasets". *Communications of the ACM*. 12(December 2021): 86–92. DOI: 10.1145/3458723.

Gebru, T. and É. P. Torres (2023). "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence". *First Monday*. 29(4– 1).

Ghosh, S. and A. Caliskan (2023). "'Person'== Light-skinned, western man, and sexualization of women of color: Stereotypes in stable diffusion". In: *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Gilardi, F., M. Alizadeh, and M. Kubli (2023). "ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks". *Proceedings of the National Academy of Sciences*. 120(30).

Girma, H. (2024). "Keynote address". In: *CSUN Assistive Technology Conference*. URL: https://www.youtube.com/watch?v=WW7erYVOej4.

Goodfriend, S. (2023). "How AI is intensifying Israel's bombardments of Gaza". *+972 Magazine*. URL: https://www.972mag.com/israel-gaza-drones-ai/.

Gotterbarn, D., A. Bruckman, C. Flick, K. Miller, and M. J. Wolf (2018). "ACM code of ethics: A guide for positive action". URL: https://cacm.acm.org/news/acm-code-of-ethic.

Grauman, K. *et al.* (2022). "Ego4D: Around the World in 3,000 Hours of Egocentric Video". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Gray, M. L. and S. Suri (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.

Green, B. (2021). "Data science as political action: Grounding data science in a politics of justice". *Journal of Social Computing*. 2(3): 249–265.

Grover, A., J. Song, A. Kapoor, K. Tran, A. Agarwal, E. J. Horvitz, and S. Ermon (2019). "Bias correction of learned generative models using likelihood-free importance weighting". In: *Advances in Neural Information Processing Systems*. Vol. 32.

Hacking, I. (2000). *The Social Construction of What?* Harvard University Press.

Hamidi, F., M. K. Scheuerman, and S. M. Branham (2018). "Gender recognition or gender reductionism? The social implications of embedded gender recognition systems". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

Hanna, A., R. Denton, A. Smart, and J. Smith-Loud (2020). "Towards a critical race methodology in algorithmic fairness". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 501–512.

Hanna, A. and T. M. Park (2020). "Against scale: Provocations and resistances to scale thinking". *ArXiv abs/2010.08850*.

Hao, K. (2022). "The biggest threat of deepfakes isn't the deepfakes themselves". *MIT Technology Review*.

Hao, K. and H. Swart (2022). "South Africa's private surveillance machine is fueling a digital apartheid". *MIT Technology Review*. URL: https://www.technologyreview.com/2022/04/19/1049996/south-africa-ai-surveillance-digital-apartheid/.

Hao, S., P. Kumar, S. Laszlo, S. Poddar, B. Radharapu, and R. Shelby (2023). "Safety and fairness for content moderation in generative models". In: *CVPR Workshop on Ethical Considerations in Creative Applications of Computer Vision*.

Hara, K., A. Adams, K. Milland, S. Savage, C. Callison-Burch, and J. 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 (CHI '18)*. New York, NY, USA, Paper 449: Association for Computing Machinery. 1–14. DOI: 10.1145/3173574.3174023.

Haraway, D. (1988). "Situated knowledges: The science question in feminism and the privilege of partial perspective". *Feminist Studies*. 14(3).

Harper, S., N. B. King, S. C. Meersman, M. E. Reichman, N. Breen, and J. Lynch (2010). "Implicit value judgments in the measurement of health inequalities". *Milbank Quarterly*. 88(1): 4–29.

Harvey, A. and J. LaPlace (2021). "Brainwash dataset". *Exposing.ai*. URL: https://exposing.ai/brainwash/.

Hecht, B., L. Wilcox, J. P. Bigham, J. Schöning, E. Hoque, J. Ernst, Y. Bisk, L. D. Russis, L. Yarosh, B. Anjum, D. Contractor, and C. Wu (2018). "It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process". *ACM Future of Computing Academy Blog*.

Heger, A. K., L. B. Marquis, M. Vorvoreanu, H. Wallach, and J. W. Vaughan (2022). "Understanding machine learning practitioners' data documentation perceptions, needs, challenges, and desiderata". *Proceedings of the ACM on Human-Computer Interaction 6, CSCW2, Article 340 (November 2022)*: 29. DOI: 10.1145/3555760.

Helmreich, S. (2000). *Silicon Second Nature: Culturing Artificial Life in a Digital World, Updated With a New Preface*. University of California Press.

Henderson, P., J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau (2020). "Towards the systematic reporting of the energy and carbon footprints of machine learning". *Journal of Machine Learning Research*. 1(Article 248): 43.

Hendricks, L. A., K. Burns, K. Saenko, T. Darrell, and A. Rohrbach (2018). "Women also snowboard: Overcoming bias in captioning models". In: *Proceedings of the European Conference on Computer vision (ECCV)*. 771–787.

Holstein, K., J. W. Vaughan, H. Daumé, M. Dudik, and H. Wallach (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 (CHI '19)*. New York, NY, USA, Paper 600: Association for Computing Machinery. 1–16. DOI: 10.1145/3290605.3300830.

Hunter, T. (2022). "Surveillance will follow us into 'the metaverse,' and our bodies could be its new data source". Washington Post. URL: https://www.washingtonpost.com/technology/2022/01/13/privacy-vr-metaverse/.

Hutchinson, B., A. Smart, A. Hanna, E. Denton, C. Greer, O. Kjartansson, P. Barnes, and M. Mitchell (2021). "Towards accountability for machine learning datasets: Practices from Software engineering and infrastructure". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. New York, NY, USA: Association for Computing Machinery. 560–575. DOI: 10.1145/3442188.3445918.

Ide, Y. and L. Beddoe (2023). "Challenging perspectives: Reflexivity as a critical approach to qualitative social work research". *Qualitative Social Work*.

Irani, L. C. and M. S. Silberman (2013). "Turkopticon: Interrupting worker invisibility in Amazon mechanical turk". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. New York, NY, USA: Association for Computing Machinery. 611–620. DOI: 10.1145/2470654.2470742.

Jacobs, A. Z. and H. Wallach (2021). "Measurement and fairness". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. New York, NY, USA: Association for Computing Machinery. 375–385. DOI: 10.1145/3442188.3445901.

Janzen, S., C. Orr, and S.-J. Terp (2022). "Cognitive security and resilience: A social ecological model of disinformation and other harms with applications to COVID-19 vaccine information behaviors". In: *ROMCIR@ ECIR*. 48–88.

Jha, A., V. Prabhakaran, R. Denton, S. Laszlo, S. Dave, R. Qadri, C. K. Reddy, and S. Dev (2024). "Beyond the surface: A global-scale analysis of visual stereotypes in text-to-image generation". *arXiv preprint arXiv:2401.06310*.

Jiang, H. H., L. Brown, J. Cheng, M. Khan, A. Gupta, D. Workman, A. Hanna, J. Flowers, and T. Gebru (2023). "AI art and its impact on artists". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 363–374.

Jo, E. S. and T. Gebru (2020). "Lessons from archives: Strategies for collecting sociocultural data in machine learning". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. New York, NY, USA: Association for Computing Machinery. 306–316. DOI: 10.1145/3351095.3372829.

Johnson, K. (2022). "How wrongful arrests based on AI derailed 3 men's lives". *Wired*. URL: https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/.

Kak, A. ed. (2020). *Regulating Biometrics: Global Approaches and Urgent Questions*. AI Now Institute.

Kalluri, P. R., W. Agnew, M. Cheng, K. Owens, L. Soldaini, and A. Birhane (2023). "The surveillance AI pipeline". *arXiv preprint arXiv:2309.15084*.

Karkkainen, K. and J. Joo (2021). "Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1548–1558.

Katzman, J., A. Wang, M. Scheuerman, S. L. Blodgett, K. Laird, H. Wallach, and S. Barocas (2023). "Taxonomizing and measuring representational harms: A look at image tagging". *Proceedings of the AAAI Conference on Artificial Intelligence*. 37(12): 14277–14285.

Kaurin, D. (2019). "Data protection and digital agency for refugees". In: *World Refugee Council Research Paper No. 12*.

Keller, H. (1934). "A neglected treasure". *The Home Magazine*.

Kermode, L., J. Freyberg, A. Akturk, R. Trafford, D. Kochetkov, R. Pardinas, E. Weizman, and J. Cornebise (2019). "Objects of violence: Synthetic data for practical ML in human rights investigations". In: *NeurIPS Joint Workshop on AI for Social Good at the 33rd Conference on Neural Information Processing Systems*.

Khan, M. and A. Hanna (2023). "The subjects and stages of AI dataset development: A framework for dataset accountability". *OSTLJ Ohio State Technology Law Journal*. 19: 171.

Khan, Z. and Y. Fu (2021). "One label, one billion faces: Usage and consistency of racial categories in computer vision". In: *Proceedings of the 2021 acm Conference on Fairness, Accountability, and Transparency*. 587–597.

Klumbytė, G., C. Draude, and A. S. Taylor (2022). "Critical tools for machine learning: Working with intersectional critical concepts in machine learning systems design". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery.

Konecný, J., H. B. McMahan, D. Ramage, and P. Richtárik (2016). "Federated optimization: Distributed machine learning for on-device intelligence". *arXiv preprint arXiv:1610.02527*.

Kovacs, A. (2020). "When our bodies become data, where does that leave us?" *Internet Democracy Project*. URL: https://internetdemocracy.in/reports/when-our-bodies-become-data/.

Krishnakumar, A., V. Prabhu, S. Sudhakar, and J. Hoffman (2021). "UDIS: Unsupervised discovery of bias in deep visual recognition models". In: *BMVC*. 143.

Latonero, M. (2019). "Stop surveillance humanitarianism". The New York Times. URL: http://web.archive.org/web/20240404082332/https://www.nytimes.com/2019/07/11/opinion/data-humanitarian-aid.html.

Latonero, M., K. Hiatt, A. Napolitano, G. Clericetti, and M. Penagos (2019). "Digital identity in the migration and refugee context: Italy case study". *Data and Society*.

Lee, S., Z. J. Wang, J. Hoffman, and D. H. P. Chau (2022). "VisCUIT: Visual auditor for bias in CNN image classifier". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 21475–21483.

Lin, H.-T., M.-F. Balcan, R. Hadsell, and M. A. Ranzato (2020). "Getting started with NeurIPS 2020". *NeurIPS Blog*. URL: https://neuripsconf.medium.com/getting-started-with-neurips-2020-e350f9b39c28.

Liu, L. T., M. Simchowitz, and M. Hardt (2019). "The implicit fairness criterion of unconstrained learning". In: *Proceedings of the 36th International Conference on Machine Learning*.

Loewenstein, A. (2023). "The Palestine laboratory: How Israel exports the technology of occupation around the world". *Verso*.

Lottick, K., S. Susai, S. Friedler, and J. Wilson (2019). "Energy usage reports: Environmental awareness as part of algorithmic accountability". In: *Workshop on Tackling Climate Change with Machine Learning at the 33rd Conference on Neural Information Processing Systems*.

Loucaides, D. (2024). "The changing face of protest mass protests used to offer a degree of safety in numbers. Facial recognition technology changes the equation". *Rest of World*. URL: https://restofworld.org/2024/facial-recognition-government-protest-surveillance/.

Luccioni, A. S., Y. Jernite, and E. Strubell (2024a). "Power hungry processing: Watts driving the cost of AI deployment?" In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 85–99.

Luccioni, S., C. Akiki, M. Mitchell, and Y. Jernite (2024b). "Stable bias: Evaluating societal representations in diffusion models". *Advances in Neural Information Processing Systems*. 36.

Mack, A., R. Qadri, R. Denton, S. Kane, and C. Bennett (2024). "'They only care to show us the wheelchair': Disability representation in text-to-image AI models". In: *CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems*.

Madaio, M. A., L. Stark, J. W. Vaughan, and H. Wallach (2020). "Co-designing checklists to understand organizational challenges and opportunities around fairness in AI". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. New York, NY, USA: Association for Computing Machinery. 1–14. DOI: 10.1145/3313831.3376445.

Madaio, M., L. Egede, H. Subramonyam, J. W. Vaughan, and H. Wallach (2022). "Assessing the fairness of AI systems: AI practitioners' processes, challenges, and needs for support". *Proceedings of the ACM on Human-Computer Interaction*. 6(CSCW1): 1–26.

McCanne, M. (2020). "When polaroid workers fought apartheid". *Dissent*.

McIntyre, D. *et al.* (1995). *Health Expenditure and Finance in South Africa*. Published jointly by the Health Systems Trust and the World Bank. 39.

McLaren, Z. M., C. Ardington, and M. Leibbrandt (2014). "Distance decay and persistent health care disparities in South Africa". *BMC Health Services Research*. 14: 1–9.

Merler, M., N. Ratha, R. S. Feris, and J. R. Smith (2019). "Diversity in faces". *arXiv preprint arXiv:1901.10436*.

Miceli, M. (2023). "Whose truth?: Power, labor, and the production of ground-truth data". *Technische Universität Berlin*. URL: https://books.google.com/books?id=734f0AEACAAJ.

Miceli, M. and J. Posada (2022). "The data-production dispositif". *Proceedings of the ACM on Human-Computer Interaction 6, CSCW2, Article 460 (November 2022)*: 37. DOI: 10.1145/3555561.

Miceli, M., T. Yang, A. A. Garcia, J. Posada, S. M. Wang, M. Pohl, and A. Hanna (2022). "Documenting data production processes: A participatory approach for data work". *Proceedings of the ACM on Human-Computer Interaction 6, CSCW2, Article 510 (November 2022)*: 34. DOI: 10.1145/3555623.

Mijente Immigrant Defense Project, and The National Immigration Project of the National Lawyers Guild (2018). "Who's behind ICE? The tech and data companies fueling deportations". URL: https://mijente.net/wp-content/uploads/2018/10/WHO%E2%80%99S-BEHIND-ICE_-The-Tech-and-Data-Companies-Fueling-Deportations-_v1.pdf.

Mijente, Just Futures Law, and No Border Wall Coalition (2021). "The deadly digital border wall". URL: https://notechforice.com/wp-content/uploads/2021/10/Deadly.Digital.Border.Wall_.pdf.

Mitchell, M., S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru (2019). "Model cards for model reporting". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. New York, NY, USA: Association for Computing Machinery. 220–229. DOI: 10.1145/3287560.3287596.

Molnar, P. (2020). *Borders and Pandemics: Surveillance Won't Stop the Coronavirus (The Ethics of COVID)*. Centre for Ethics, University of Toronto. URL: https://www.youtube.com/watch?v=4sMAF4fnK3g.

Molnar, P. (2024). "The walls have eyes : Surviving migration in the age of artificial intelligence". *New Press*.

Mutembesa, D., E. Mwebaze, S. Nsumba, C. Omongo, and H. Mutaasa (2019). "Mobile community sensing with smallholder farmers in a developing nation; A scaled pilot for crop health monitoring". *arXiv preprint arXiv:1908.07047*.

Nakate, V. (2021). ""A Bigger Picture": Ugandan Activist Vanessa Nakate on Bringing New Voices to the Climate Fight". Democracy Now. URL: https://www.democracynow.org/2021/12/13/a_bigger_picture_by_vanessa_nakate.

Narayanan, A. (2018). "21 fairness definitions and their politics". In: *ACM Conference on Fairness, Accountability, and Transparency. Tutorials*. URL: https://www.youtube.com/watch?v=jIXIuYdnyyk.

NeurIPS (2022). "Paper checklist guidelines". URL: https://neurips.cc/Conferences/2022/PaperInformation/PaperChecklist.

NeurIPS (2023). "Datasets and benchmarks track". URL: https://neurips.cc/Conferences/2023/CallForDatasetsBenchmarks.

Nicoletti, L. and D. Bass (2023). "Humans are biased. Generative AI is even worse". *Bloomberg*. URL: https://www.bloomberg.com/graphics/2023-generative-ai-bias/.

Nightingale, S. J. and H. Farid (2022). "AI-synthesized faces are indistinguishable from real faces and more trustworthy". *Proceedings of the National Academy of Sciences*. 119(8): e2120481119.

Noble, S. U. (2018). "Algorithms of oppression: How search engines reinforce racism". In: *Algorithms of Oppression*. New York University Press.

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Oreskes, N. (2022). *Science on a Mission: How Military Funding Shaped What We Do and Don't Know About the Ocean*. University of Chicago Press.

Parraga, O., M. D. More, C. M. Oliveira, N. S. Gavenski, L. S. Kupssinskü, A. Medronha, L. V. Moura, G. S. Simões, and R. C. Barros (2023). "Debiasing methods for fairer neural models in vision and language research: A survey". *ACM Computing Surveys*.

Partnership on AI (2021). "Managing the risks of AI research: Six recommendations for responsible publication". URL: https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Managing-the-Risks-of-AI-Resesarch-Responsible-Publication.pdf.

Passi, S. and S. Barocas (2019). "Problem formulation and fairness". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. New York, NY, USA: Association for Computing Machinery. 39–48. DOI: 10.1145/3287560.3287567.

Paullada, A., I. D. Raji, E. M. Bender, E. Denton, and A. Hanna (2021). "Data and its (dis) contents: A survey of dataset development and use in machine learning research". *Patterns*. 2(11).

Peng, K. L., A. Mathur, and A. 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)*.

Perrigo, B. (2022). "Inside facebook's African sweatshop". *TIME Magazine*. URL: https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.

Piergiovanni, A. J. and M. Ryoo (2020). "Avid dataset: Anonymized videos from diverse countries". *Advances in Neural Information Processing Systems*. 33: 16711–16721.

Poulson, J. (2022). "Militaries, intelligence agencies, and law enforcement dominate U.S. and U.K. government purchasing from U.S. tech giants". *Tech Inquiry*. URL: https://techinquiry.org/docs/InternationalCloud.pdf.

Prabhu, V., R. R. Selvaraju, J. Hoffman, and N. Naik (2022). "Can domain adaptation make object recognition work for everyone?" In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3981–3988.

Prescod-Weinstein, C. (2021). *The Disordered Cosmos: A Journey into Dark Matter, Spacetime, and Dreams Deferred*. Bold Type Books.

Prescod-Weinstein, C. (2022). "Even underwater, money talks". *Physics Today*. 75(7): 53–54.

Prunkl, C., C. Ashurst, M. Anderljung, H. Webb, J. Leike, and A. Dafoe (2021). "Institutionalizing ethics in AI through broader impact requirements". *Nature Machine Intelligence*. 3: 104–110.

Pushkarna, M., A. Zaldivar, and O. Kjartansson (2022). "Data cards: Purposeful and transparent dataset documentation for responsible AI". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 1776–1826. DOI: 10.1145/3531146.3533231.

Qadri, R., R. Shelby, C. L. Bennett, and R. Denton (2023). "Ai's regimes of representation: A community-centered study of text-to-image models in south Asia". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 506–517.

Raji, D., E. Denton, E. M. Bender, A. Hanna, and A. Paullada (2021). "AI and the everything in the whole wide world benchmark". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*.

Raji, I. D., T. Gebru, M. Mitchell, J. Buolamwini, J. Lee, and E. Denton (2020). "Saving face: Investigating the ethical concerns of facial recognition auditing". In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 145–151.

Ramaswamy, V. V., S. Y. Lin, D. Zhao, A. Adcock, L. van der Maaten, D. Ghadiyaram, and O. Russakovsky (2023). "Geode: A geographically diverse evaluation dataset for object recognition". In: *Advances in Neural Information Processing Systems*. Vol. 36.

Ranganathan, N. (2020). "The economy (and regulatory practice) that biometrics inspires: A study of the aadhaar project". In: *AI Now Report on Regulating Biometrics: Global Approaches and Urgent Questions*.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi (2016). "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 779–788.

Ren, Z., Y. J. Lee, and M. S. Ryoo (2018). "Learning to anonymize faces for privacy preserving action detection". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 620–636.

Rhue, L. (2018). "Racial influence on automated perceptions of emotions". *SSRN Electronic Journal*.

Roberts, D. E. (2017). "Democratizing criminal law as an abolitionist project, 111 Nw. U. L. Rev. 1597". URL: https://scholarlycommons.law.northwestern.edu/nulr/vol111/iss6/11.

Robinette, P., W. Li, R. Allen, A. M. Howard, and A. R. Wagner (2016). "Overtrust of robots in emergency evacuation scenarios". In: *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

Rogaway, P. (2015). "The moral character of cryptographic work". *Cryptology ePrint Archive*.

Rostamzadeh, N., D. Mincu, S. Roy, A. Smart, L. Wilcox, M. Pushkarna, J. Schrouff, R. Amironesei, N. Moorosi, and K. Heller (2022). "Healthsheet: Development of a transparency artifact for health datasets". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York, NY, USA: Association for Computing Machinery. 1943–1961. DOI: 10.1145/3531146.3533239.

Ryu, H. J., H. Adam, and M. Mitchell (2018). "Inclusive facenet: Improving face attribute detection with race and gender diversity". In: *Proceedings of FAT/ML*.

Sambasivan, N., E. Arnesen, B. Hutchinson, T. Doshi, and V. Prabhakaran (2021a). "Re-imagining algorithmic fairness in india and beyond". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 315–328.

Sambasivan, N., S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo (2021b). "Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Vol. Article 39. New York, NY, USA: Association for Computing Machinery. 1–15. DOI: 10.1145/3411764.3445518.

Scheuerman, M. K., A. Hanna, and E. Denton (2021). "Do datasets have politics? Disciplinary values in computer vision dataset development". *Proceedings of the ACM on Human-Computer Interaction 5, CSCW2, Article 317 (October 2021)*: 37. DOI: 10.1145/3476058.

Schumann, C., F. Olanubi, A. Wright, E. Monk, C. Heldreth, and S. Ricco (2024). "Consensus and subjectivity of skin tone annotation for ml fairness". In: *Advances in Neural Information Processing Systems*. Vol. 36.

Schumann, C., S. Ricco, U. Prabhu, V. Ferrari, and C. Pantofaru (2021). "A step toward more inclusive people annotations for fairness". In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 916–925.

Schwartz, R., J. Dodge, N. A. Smith, and O. Etzioni (2020). "Green AI". *Communications of the ACM*. 12(December 2020): 54–63. DOI: 10.1145/3381831.

Sefala, R., T. Gebru, L. Mfupe, N. Moorosi, and R. Klein (2021). "Constructing a visual dataset to study the effects of spatial apartheid in South Africa". In: *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Selbst, A. D., D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi (2019). "Fairness and abstraction in sociotechnical systems". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. New York, NY, USA: Association for Computing Machinery. 59–68. DOI: 10.1145/3287560.3287598.

Shah, N. (2015). "Identity and identification: The individual in the time of networked governance". *Socio Legal Review*. 11(2): 22.

Shan, S., J. Cryan, E. Wenger, H. Zheng, R. Hanocka, and B. Y. Zhao (2023). "Glaze: Protecting artists from style mimicry by {Text-to-Image} models". In: *32nd USENIX Security Symposium (USENIX Security 23)*. 2187–2204.

Shelby, R., S. Rismani, K. Henne, A. J. Moon, N. Rostamzadeh, P. Nicholas, N. Yilla-Akbari, *et al.* (2023). "Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction". In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 723–741.

Sherman, J. (2021). "Data brokers are a threat to democracy". *Wired*. URL: https://www.wired.com/story/opinion-data-brokers-are-a-threat-to-democracy/.

Sjoding, M. W., R. P. Dickson, T. J. Iwashyna, S. E. Gay, and T. S. Valley (2020). "Racial bias in pulse oximetry measurement". *New England Journal of Medicine*. 383(25): 2477–2478.

Solaiman, I., Z. Talat, W. Agnew, L. Ahmad, D. Baker, S. L. Blodgett, H. Daumé III, J. Dodge, E. Evans, S. Hooker, Y. Jernite, A. S. Luccioni, A. Lusoli, M. Mitchell, J. Newman, M.-T. Png, A. Strait, and A. Vassilev (2023). "Evaluating the social impact of generative AI systems in systems and society". *ArXiv abs/2306.05949*.

Solomon, B. (2018). "Digital IDs are more dangerous than you think". *Wired*. URL: https://www.wired.com/story/digital-ids-are-more-dangerous-than-you-think/.

Srinivasan, R., E. Denton, J. Famularo, N. Rostamzadeh, F. Diaz, and B. Coleman (2021). "Artsheets for art datasets". *NeurIPS Datasets and Benchmarks Track*.

Stark, L. (2019). "Facial recognition is the plutonium of AI". *XRDS 25*. 3(Spring 2019): 50–55. DOI: 10.1145/3313129.

Stark, L. and J. Hutson (2021). "Physiognomic artificial intelligence". *Fordham Intellectual Property, Media and Entertainment Law Journal*. 32: 922.

Stewart, R., M. Andriluka, and A. Y. Ng (2016). "End-to-end people detection in crowded scenes". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2325–2333.

Stitzlein, S. M. (2004). "Replacing the 'View from Nowhere': A pragmatist-feminist science classroom". *Electronic Journal of Science Education*. 9(2).

Stoyanovich, J. and B. Howe (2019). "Nutritional labels for data and models". *A Quarterly Bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*.

Strubell, E., A. Ganesh, and A. McCallum (2019). "Energy and policy considerations for deep learning in NLP". *Proceedings of the AAAI Conference on Artificial Intelligence*. 34(09): 13693–13696.

Suresh, H. and J. Guttag (2021). "A framework for understanding sources of harm throughout the machine learning life cycle". In: *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–9.

Sweeney, L. (2013). "Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising". *Queue*. 11(3): 10–29.

Taylor, K.-Y. (2019). *Race for Profit: How Banks and the Real Estate Industry Undermined Black Homeownership*. The University of North Carolina Press.

The Distributed AI Research Institute (DAIR) (2022). "DAIR research philosophy". URL: https://www.dair-institute.org/research-philosophy/.

The Engine Room and Oxfam (2018). "Biometrics in the Humanitarian Sector". URL: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620454/rr-biometrics-humanitarian-sector-050418-en.pdf?sequence=1.

The People + AI Guidebook. (2019). URL: http://pair.withgoogle.com/guidebook.

Trikanad, S. and V. Bhandari (2022). "Surveillance enabling identity systems in Africa: Tracing the fingerprints of aadhaar (August 6, 2022). Centre for Internet and Society". URL: https://ssrn.com/abstract=4239604.

Turkopticon (2023). "Beware the Hype: ChatGPT didn't replace human data annotators". *Tech Workers Coalition Newsletter*. (5). URL: https://news.techworkerscoalition.org/2023/04/04/issue-5/.

UN High Commissioner for Refugees (UNHCR) (2022). *Climate Change, displacement and human rights*. URL: https://www.refworld.org/reference/themreport/unhcr/2022/en/124057.

Ungless, E. L., B. Ross, and A. Lauscher (2023). "Stereotypes and smut: The (Mis) representation of non-cisgender identities by text-to-image models". In: *The 61st Annual Meeting of the Association for Computational Linguistics*.

United Nations Department of Economic and Social Affairs (2011). *The global social crisis: Report of the World social situation 2011*. URL: https://www.un-ilibrary.org/content/books/9789210552226/read.

Venkatasubramanian, S., T. Gebru, U. Topcu, H. Griffin, L. Rosenbloom, and N. Sonboli (2024). *Community driven approaches to research in technology and society CCC workshop report*. URL: https://cra.org/ccc/wp-content/uploads/sites/2/2024/03/CDARTS-Workshop-Report_Final.pdf.

Venturini, J. and V. Garay (2022). "Facial recognition in Latin America: Trends in the implementation of a perverse technology". Al Sur. URL: https://creativecommons.org/licenses/by/4.0/legalcode.es.

Villeneuve, S., M. Andrus, and H. Hongo (2021). "Shedding light on the trade-offs of using demographic data for algorithmic fairness". *Partnership on AI Blog*.

Vince, G. (2022). "Is the world ready for mass migration due to climate change?" *BBC*. URL: https://www.bbc.com/future/article/20221117-how-borders-might-change-to-cope-with-climate-migration.

Vincent, J. (2021). "Facebook is researching AI systems that see, hear, and remember everything you do". *The Verge*. URL: https://www.theverge.com/2021/10/14/22725894/facebook-augmented-reality-ar-glasses-ai-systems-ego4d-research.

Wagstaff, K. L. (2012). "Machine learning that matters". In: *Proceedings of the 29th International Coference on International Conference on Machine Learning*. 1851–1856.

Wan, Y., A. Subramonian, A. Ovalle, Z. Lin, A. Suvarna, C. Chance, H. Bansal, R. Pattichis, and K.-W. Chang (2024). "Survey of bias in text-to-image generation: Definition, evaluation, and mitigation". *arXiv preprint arXiv:2404.01030*.

Wang, A., S. Barocas, K. Laird, and H. Wallach (2022a). "Measuring representational harms in image captioning". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 324–335.

Wang, A., A. Liu, R. Zhang, A. Kleiman, L. Kim, D. Zhao, I. Shirai, A. Narayanan, and O. Russakovsky (2022b). "REVISE: A tool for measuring and mitigating bias in visual datasets". *International Journal of Computer Vision*. 130(7): 1790–1810.

Wang, D., S. Prabhat, and N. Sambasivan (2022c). "Whose AI dream? In search of the aspiration in data annotation". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA. Article 582. 1–16. DOI: 10.1145/3491102.3502121.

Wang, M., W. Deng, J. Hu, X. Tao, and Y. Huang (2019a). "Racial faces in the wild: Reducing racial bias by information maximization adaptation network". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 692–702.

Wang, T., J. Zhao, M. Yatskar, K.-W. Chang, and V. Ordonez (2019b). "Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5310–5319.

Wang, Z., K. Qinami, I. C. Karakozis, K. Genova, P. Nair, K. Hata, and O. Russakovsky (2020). "Towards fairness in visual recognition: Effective strategies for bias mitigation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8919–8928.

Warden, P., M. Stewart, B. Plancher, S. Katti, and V. J. Reddi (2023). "Machine learning sensors". *Communications of the ACM*. 66(11): 25–28.

Weerts, H., M. Dudík, R. Edgar, A. Jalali, R. Lutz, and M. Madaio (2023). "Fairlearn: Assessing and improving fairness of AI systems". *Journal of Machine Learning Research*. 24: 1–8.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman and Co.

West, S. M., M. Whittaker, and K. Crawford (2019). "Discriminating systems". *AI Now*. (2019): 1–33.

Williams, A. (2021). *Surveillance Nation: The Real Cost of Amazon's Customer Obsession*. The Distributed AI Research Institute (DAIR). URL: https://peertube.dair-institute.org/w/oAWnan9PXTvS3CyutatQcg.

Williams, A., M. Miceli, and T. Gebru (2022). "The exploited labor behind artificial intelligence". *Noema Magazine*. URL: https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/.

Wilson, B., J. Hoffman, and J. Morgenstern (2019). "Predictive inequity in object detection". *arXiv preprint arXiv:1902.11097*.

Wolfe, R., Y. Yang, B. Howe, and A. Caliskan (2023). "Contrastive language-vision AI models pretrained on web-scraped multimodal data exhibit sexual objectification bias". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1174–1185.

Wong, R. Y., K. Boyd, J. Metcalf, and K. Shilton (2020). "Beyond checklist approaches to ethics in design". In: *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '20 Companion). Association for Computing Machinery*. New York, NY, USA. 511–517. DOI: 10.1145/3406865.3418590.

Wu, C. J., R. Raghavendra, U. Gupta, B. Acun, N. Ardalani, K. Maeng, G. Chang, F. Aga, J. Huang, C. Bai, and M. Gschwind (2022). "Sustainable AI: Environmental implications, challenges and opportunities". *Proceedings of Machine Learning and Systems*. 4: 795–813.

Yang, K., K. Qinami, L. Fei-Fei, J. Deng, and O. Russakovsky (2020). "Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 547–558.

Yang, K., J. H. Yau, L. Fei-Fei, J. Deng, and O. Russakovsky (2022). "A study of face obfuscation in imagenet". In: *International Conference on Machine Learning*. PMLR. 25313–25330.

Yew, R.-J., L. Qin, and S. Venkatasubramanian (2023). "You still see me: How data protection supports the architecture of ML surveillance". In: *NeurIPS, 2023 Workshop on Regulatable ML*.

York, G. and Z. Zelalem (2024). "Cheap imported drones fuel a surge of deaths in Africa". *The Globe and Mail*. URL: https://www.theglobeandmail.com/world/article-cheap-imported-drones-fuel-a-surge-of-deaths-in-africa.

Zelalem, Z. (2021). "Refugee camps in ethiopia appear to have been systematically destroyed". *Vice News*. URL: https://www.vice.com/en/article/93wmbz/refugee-camps-in-ethiopia-appear-to-have-been-systematically-destroyed.

Zemel, R., Y. Wu, K. Swersky, T. Pitassi, and C. Dwork (2013). "Learning fair representations". In: *International Conference on Machine Learning*. PMLR. 325–333.

Zhang, B. H., B. Lemoine, and M. Mitchell (2018). "Mitigating unwanted biases with adversarial learning". In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 335–340.

Zhao, J., T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang (2017). "Men also like shopping: Reducing gender bias amplification using corpus-level constraints". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

Zuboff, S. (2018). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*. Profile Books.