Article

Evaluating Dataset Bias in Biometric Research

Betty Heleen *

Posted Date: 20 August 2025

doi: 10.20944/preprints202508.1479.v1

Keywords: biometric systems; dataset bias; fairness; inclusivity; demographic representation; ethical AI; bias mitigation; biometric research; algorithmic discrimination; human-centered design

Article

Evaluating Dataset Bias in Biometric Research

Betty Heleen

Independent Researcher, North Macedonia; bettyheleen108@gmail.com

Abstract

Biometric technologies are rapidly becoming central to identity verification, national security, and personalized services. However, as the reliance on biometric systems grows, so do the concerns around dataset bias, a silent but significant threat to fairness, accuracy, and inclusivity. This paper investigates the presence and impact of dataset bias in biometric research, shedding light on how skewed data representation can lead to discriminatory outcomes, particularly for underrepresented groups. We explore the roots of bias, ranging from limited demographic diversity in training datasets to socio-technical factors influencing data collection practices. Through real-world case studies and critical analysis, this study urges researchers and developers to adopt more ethical, transparent, and inclusive data strategies. The goal is not just to improve biometric system performance but to ensure that these technologies serve all individuals equally, regardless of race, gender, or age. Tackling dataset bias isn't just a technical issue; it's a matter of social justice and trust in an increasingly digital world.

Keywords: biometric systems; dataset bias; fairness; inclusivity; demographic representation; ethical AI; bias mitigation; biometric research; algorithmic discrimination; human-centered design

1. Introduction

1.1. Background and Context

Biometric technologies, once confined to science fiction, are now embedded in our daily lives, from unlocking smartphones to verifying identities at border controls. These systems rely on biological and behavioral characteristics such as fingerprints, facial features, iris patterns, and voiceprints to authenticate individuals. As adoption spreads across sectors finance, healthcare, law enforcement, and beyond, biometrics are increasingly viewed as not only efficient but also more secure than traditional identification methods. However, beneath the technical advancements lies a quieter but critical concern: the integrity of the datasets used to train and evaluate these systems. Biometric systems are only as unbiased and effective as the data they learn from. When that data is skewed, overrepresenting certain demographics while neglecting others, it sets the stage for uneven performance. This issue of dataset bias is not theoretical. It has been observed, measured, and experienced in real-world scenarios, where individuals from underrepresented groups encounter higher error rates, misidentifications, or even systemic exclusion. Such discrepancies challenge the fairness, accuracy, and trustworthiness of biometric technologies.

1.2. Significance of Dataset Bias in Biometrics

Bias in biometric datasets is more than a technical glitch; it is a social and ethical fault line. As societies increasingly automate identity verification and surveillance, decisions once made by humans are being delegated to machines trained on historical data. If that data reflects or amplifies existing societal inequities, the technology can perpetuate discrimination in new, opaque ways. In law enforcement, for instance, a biased facial recognition system can disproportionately misidentify individuals from certain racial backgrounds, leading to wrongful arrests or surveillance. In healthcare, faulty biometric verification might limit access to essential services. In such contexts, dataset bias can erode public trust, exacerbate existing inequalities, and trigger legal and ethical concerns. Addressing this issue is therefore not just about improving system accuracy; it's about safeguarding human dignity and rights.

1.3. Objectives and Scope of the Study

This article aims to provide a critical, in-depth evaluation of dataset bias in biometric research. It explores the different types and sources of bias, analyzes their consequences on system performance and human lives, and reviews real-world case studies that illustrate the severity of the problem. Furthermore, it investigates current methods for detecting and measuring bias, and proposes strategies, both technical and policy-driven, for mitigation. While much of the existing literature focuses on facial recognition, this study extends the lens to other modalities such as fingerprint and iris recognition, offering a more holistic view of the landscape. The goal is not only to highlight the challenges but to contribute meaningfully to the ongoing conversation about ethical, inclusive, and accountable biometric systems.

2. Understanding Dataset Bias

2.1. Definition and Types of Bias

Dataset bias refers to systematic errors or imbalances in data that influence the outcomes of machine learning models. In the context of biometrics, this means that the training data does not adequately represent the diversity of the population it intends to serve. This can lead to unequal performance across demographic groups what works well for one group might work poorly for another.

Several types of dataset bias have been identified in biometric research. **Sampling bias** occurs when certain populations are over- or underrepresented in the dataset, leading to skewed performance metrics. **Measurement bias** arises from inconsistencies in how biometric data is captured, for instance, varying lighting conditions or sensor types. **Label bias** may occur when ground truth labels (such as identity tags) are inaccurate or inconsistently applied. These biases, whether unintentional or systemic, can seep into model training and manifest as errors during real-world deployment.

2.2. Sources of Bias in Biometric Data Collection

Bias can creep in at multiple stages of data collection. One common source is **reliance on convenience samples**, such as university students, corporate volunteers, or specific geographic regions. While these datasets may be easy to gather, they rarely reflect global diversity in terms of age, ethnicity, gender, or cultural backgrounds. Another source is **technological limitations**. Biometric devices cameras, fingerprint sensors, or iris scanners, may perform differently based on skin tone, age, or physical conditions. For instance, facial recognition systems often perform worse on individuals with darker skin due to insufficient infrared reflectance or poor contrast in image capture. Even the **design of collection protocols** can introduce bias. Are participants equally instructed? Is the environment controlled across all demographics? Seemingly minor decisions like lighting or camera angles can disproportionately affect certain groups and distort the dataset's representativeness.

2.3. Demographic Disparities and Underrepresentation

Demographic bias is perhaps the most widely documented form of dataset bias in biometric research. Studies consistently show that biometric systems, especially those powered by deep learning, struggle with underrepresented groups. For example, women, older adults, people with disabilities, and individuals from minority ethnic backgrounds often face higher error rates in face and voice recognition systems. This underrepresentation is not always malicious; it's often the result of data availability, commercial prioritization, or even regional constraints. Nonetheless, its impact is deeply problematic. A system trained primarily on adult Caucasian males may achieve impressive benchmark scores but fail in real-world settings that demand inclusivity.

3. Impacts of Dataset Bias on Biometric Systems

3.1. Accuracy and Performance Discrepancies

When datasets are biased, so are the systems built upon them. Biometric algorithms trained on skewed data exhibit **variable accuracy across demographic lines**. This disparity undermines the core promise of biometrics' universal, reliable identification. In facial recognition, for example, studies by MIT Media Lab and NIST (National Institute of Standards and Technology) have shown error rates for women and people of color to be several times higher than those for white males. These discrepancies are not mere statistical noise; they translate to inconsistent and unreliable user experiences. A fingerprint scanner that works flawlessly for one demographic might routinely reject valid inputs from another. Such inconsistency is especially dangerous in high-stakes settings like airport security, law enforcement, and financial authentication, where a false rejection or false acceptance can lead to serious consequences.

3.2. Real-World Consequences for Marginalized Groups

Dataset bias in biometrics doesn't just affect code or error logs; it affects lives. Consider a Black teenager misidentified by a facial recognition system and wrongfully arrested. Or an elderly person consistently locked out of a biometric-based healthcare portal because their fingerprint scans fail repeatedly. These examples are not hypothetical; they have occurred and have been documented in the media and legal records. Marginalized communities bear the brunt of these failures. The intersection of **systemic bias** and **automated decision-making** can compound social inequities. For those already disadvantaged, being misclassified, excluded, or surveilled by biometric systems can deepen mistrust in public institutions and technology at large.

3.3. Ethical and Legal Implications

The ethical dimensions of dataset bias are profound. Biometric systems often operate invisibly; people may not know when, where, or how they are being scanned. When errors occur, there may be little recourse for the affected individuals, especially if the systems lack transparency or human oversight. From a legal standpoint, the consequences are equally significant. Countries like the European Union have enacted strict regulations under the General Data Protection Regulation (GDPR), emphasizing fairness, transparency, and accountability in AI and biometric systems. Failing to mitigate dataset bias can lead not only to reputational harm but also to **legal penalties and loss of public trust**. Moreover, the lack of standardized global frameworks makes it difficult to hold developers and deployers accountable. In jurisdictions without strong legal protections, biased biometric systems can become tools of **discrimination and surveillance**, rather than empowerment.

4. Case Studies and Examples

4.1. Facial Recognition Failures

Facial recognition has become the poster child for both the promise and peril of biometric technology. Several high-profile failures have exposed the deep cracks caused by dataset bias. One of the most widely cited studies by Joy Buolamwini and Timnit Gebru at the MIT Media Lab found that commercial facial recognition systems had error rates as high as **34.7% for dark-skinned women**, compared to just **0.8% for light-skinned men**. These systems, trained predominantly on lighter-skinned male faces, struggled to accurately identify individuals outside that narrow demographic. Real-world consequences followed. In 2020, **Robert Williams**, a Black man from Michigan, was wrongfully arrested after a facial recognition system falsely matched him to security footage from a theft. He spent 30 hours in police custody before authorities admitted the mistake. This was not an isolated incident; similar cases in New Jersey and the UK have shown how biased data can lead to false accusations, particularly against communities of color. These failures illustrate the urgent need for better representation and transparency in the datasets powering facial recognition systems.

4.2. Fingerprint and Iris Recognition Bias

While facial recognition has received the most scrutiny, **fingerprint and iris recognition systems are not immune to bias**. Several studies have shown that fingerprint sensors can have **lower match rates** for individuals with darker skin tones, worn-down fingerprints (common among manual laborers), or certain medical conditions affecting the epidermis. In one study, researchers found that older adults and people with diabetes often experience higher **False Non-Match Rates (FNMR)** in fingerprint systems. Similarly, iris recognition systems may perform poorly for individuals with certain eye conditions or those from geographic regions where lighter eye pigmentation is rare. For example, some infrared-based iris scanners struggle with brown or dark irises due to **lower contrast**, leading to increased rejection rates. These limitations highlight the importance of designing both hardware and algorithms with diverse populations in mind. A "one-size-fits-all" biometric system, when trained on narrow datasets, simply does not serve everyone equally.

4.3. Regional and Cultural Representation Issues

Biometric systems deployed globally often rely on datasets developed in a few tech hubs, largely in North America, Europe, and East Asia. This leads to significant **underrepresentation of populations in South Asia, Sub-Saharan Africa, Latin America, and Indigenous communities**. The result? Systems that falter in unfamiliar cultural contexts, both technically and socially. For example, in some cultures, facial coverings, religious garments, or customary headwear can interfere with biometric capture and recognition. Systems not trained on such variations often flag these inputs as errors or anomalies. In one deployment of a biometric voting system in Kenya, **authentication failure rates were significantly higher** in rural areas compared to urban centers, largely due to poor image capture, environmental factors, and insufficiently localized training data. These regional disparities are not just technical hiccups; they represent a failure to account for the global, multicultural reality of biometric deployment.

5. Bias Detection and Measurement Techniques

5.1. Metrics and Evaluation Tools

To address dataset bias, we first need to measure it effectively. Common metrics include **False Match Rate (FMR)**, **False Non-Match Rate (FNMR)**, and **Equal Error Rate (EER),** all of which help quantify a system's performance across different groups. However, these metrics alone do not capture demographic discrepancies. Researchers increasingly use **disaggregated performance metrics**, evaluating accuracy by race, gender, age, and other demographic categories. A model may report a 95% overall accuracy, but that figure can mask wildly different outcomes across subgroups. For example, a system may yield **98% accuracy for white males but only 80% for Black females,** a hidden inequity unless metrics are stratified. In addition, techniques like **confusion matrices** and **Receiver Operating Characteristic (ROC) curves** can be applied to subgroup performance. When used alongside demographic labels, these tools can uncover hidden biases and drive more equitable system development.

5.2. Benchmark Datasets and Limitations

Benchmark datasets are widely used in biometric research to evaluate and compare algorithmic performance. Notable examples include **LFW (Labeled Faces in the Wild)**, **CASIA**, and **NIST's**

doi:10.20944/preprints202508.1479.v1

FRVT datasets. However, many of these datasets suffer from the same biases they're meant to expose. LFW, for example, is known to be dominated by celebrities, largely white and male, introducing inherent skew in facial recognition research. Even newer datasets that claim diversity often lack transparency in **demographic annotations** or **collection methodologies**. Without clear documentation of data sources, consent protocols, and representation breakdowns, these datasets cannot reliably serve as fair evaluation tools. Some researchers argue that current benchmarking practices reinforce biased results by rewarding performance on already skewed data. There is a growing call within the research community for **more representative, ethically sourced, and publicly accountable benchmark datasets** that reflect real-world demographics more faithfully.

5.3. Transparency and Auditing Practices

Transparency is the bedrock of ethical AI, yet biometric vendors and researchers often keep their datasets and model architectures under wraps. This lack of openness makes it difficult for external parties to audit systems for bias or test generalizability across diverse populations. **Algorithmic auditing,** a practice that involves reviewing both datasets and model behavior, can play a key role here. Independent audits, conducted by third-party researchers or regulatory bodies, can surface hidden issues in biometric systems and push companies toward more accountable practices. Some initiatives, such as **Datasheets for Datasets** and **Model Cards**, encourage developers to document dataset composition, intended uses, and known limitations. Though still not industry standard, these transparency tools are steps in the right direction. Without them, we risk deploying black-box systems whose errors disproportionately harm the very groups they claim to serve.

6. Strategies for Mitigating Dataset Bias

6.1. Inclusive Data Collection and Curation

Tackling dataset bias in biometric research begins at the source: the data itself. Inclusive data collection isn't just about gathering more data; it's about **gathering the right data** from a diverse and representative pool of individuals. This includes ensuring equitable representation across variables like race, age, gender, disability status, and geographic origin. Ethical data practices start with informed consent, community engagement, and sensitivity to cultural norms. For instance, involving local communities when collecting data in underrepresented regions can build trust and ensure authenticity. Researchers should also seek to **balance dataset demographics intentionally**, avoiding dominance of any single group. Moreover, post-collection **data curation** plays a key role. This means auditing the dataset for skewed patterns, correcting imbalances, and documenting every decision—what was included, excluded, and why. By making dataset composition transparent and purpose-driven, researchers can mitigate bias before the model even sees the data.

6.2. Algorithmic Fairness Techniques

Even with improved datasets, algorithms themselves can perpetuate bias. That's where **algorithmic fairness** techniques come into play. These are mathematical strategies designed to equalize outcomes across demographic groups without sacrificing performance. One common approach is **re-weighting** training data so that underrepresented groups have a proportional impact on the model's learning process. Another is **adversarial de-biasing**, where a secondary model tries to predict protected attributes (e.g., race or gender) from the primary model's output; if it succeeds, the primary model is penalized and adjusted accordingly. This technique encourages the system to focus on task-relevant features rather than biased correlations. There are also **post-processing techniques**, such as calibrating the threshold for different groups to ensure equal false-positive or false-negative rates. While these methods are not silver bullets, they serve as practical tools to correct biases that data balancing alone cannot resolve.

6.3. Policy and Governance Recommendations

Technology alone cannot solve a problem rooted in systemic inequities. Effective mitigation also requires **robust policy and governance frameworks**. Regulators, institutions, and developers must align on standards that enforce fairness, transparency, and accountability. This includes mandating **bias audits** for biometric systems before deployment, requiring **demographic performance reporting**, and establishing **independent oversight committees** to review system impacts. Legal frameworks such as the EU's **AI Act** and existing privacy laws (e.g., GDPR) provide starting points, but more specific guidelines are needed to address the nuances of biometric technologies. Governments and international bodies can also support open, diverse datasets and fund research into ethical AI practices. Without these top-down interventions, market forces alone are unlikely to prioritize inclusivity, especially when speed and cost efficiency dominate the agenda.

7. Research Gaps and Future Directions

7.1. Need for Global Collaboration

One of the most glaring gaps in current biometric research is the **lack of international collaboration**. Much of the cutting-edge work is done in silos—by Western tech companies, elite universities, or private firms with proprietary goals. This not only limits the scope of research but also risks exporting biased systems into countries and communities that had no say in their design. To create genuinely fair biometric systems, we need **cross-border, interdisciplinary partnerships** that include governments, academic institutions, civil society, and affected communities. Global initiatives should aim to establish ethical data-sharing frameworks, co-develop inclusive datasets, and ensure equitable access to tools and knowledge. Fairness in biometrics must be treated as a shared global responsibility, not just a competitive advantage.

7.2. Calls for Standardized Ethical Guidelines

Despite the increasing visibility of bias in AI, **standardized ethical guidelines** for biometric data collection and model evaluation remain largely absent. Different companies and research labs operate under vastly different assumptions about what constitutes fairness, consent, or inclusivity. There is a pressing need for international consensus on what ethical biometric development looks like. This includes standardized demographic benchmarks, bias reporting templates, and impact assessments. Without shared norms, even well-meaning efforts can result in fragmented or inconsistent safeguards. Organizations like the IEEE and ISO have begun proposing technical standards, but adoption remains uneven. To fill this gap, professional associations, governments, and funding bodies must **mandate ethical standards as part of biometric research and procurement processes**.

7.3. Opportunities for Interdisciplinary Research

Finally, the future of fair biometrics lies not only in computer science but in **interdisciplinary collaboration**. Understanding the societal impacts of dataset bias requires perspectives from sociology, law, ethics, anthropology, and public policy. Technical fixes can only go so far without a grounded understanding of how biometric systems interact with human identities and power structures. For example, sociologists can help uncover how biometric errors disproportionately affect marginalized communities. Legal scholars can interpret how new data protection laws reshape deployment norms. Anthropologists can offer insight into how cultural practices influence biometric usability. Bridging these disciplines opens up new avenues of inquiry, richer datasets, better-informed algorithms, and more equitable outcomes. The challenge is complex, but the path forward is clear: **we need to stop building systems in isolation and start designing them in conversation**.

8. Conclusions

8.1. Summary of Findings

Dataset bias in biometric research is not a peripheral issue; it is central to the technology's accuracy, fairness, and legitimacy. This article has explored how bias originates from imbalanced data collection, technological limitations, and systemic underrepresentation, leading to disparate outcomes across demographic groups. We've seen how biased biometric systems can perform inconsistently, disproportionately misidentify marginalized populations, and trigger real-world harm, ranging from wrongful arrests to service denial. Case studies in facial, fingerprint, and iris recognition reveal that these issues are not limited to academic debate; they are already affecting lives, often in invisible and irreversible ways. Current evaluation tools and benchmark datasets, though widely used, often fall short in detecting and correcting for such biases. Despite ongoing efforts, the lack of inclusive data practices, transparency, and ethical governance continues to threaten the equitable deployment of biometric systems worldwide.

8.2. The Road Ahead for Ethical Biometrics

Solving the problem of dataset bias is both a technical and moral imperative. Inclusive data collection, fairness-driven algorithms, and regulatory oversight must be pursued in parallel, not as optional enhancements, but as foundational pillars of biometric system design. These changes require a cultural shift in how we define success in AI, not just by accuracy or speed, but by equity, inclusivity, and respect for human dignity. The road ahead is collaborative. Researchers, developers, regulators, and communities must work together across borders and disciplines to build biometric systems that serve everyone, not just the statistically dominant or technologically convenient. This will involve challenging the status quo, embracing transparency, and investing in long-term solutions that may not offer quick wins but promise lasting impact. In the end, the goal is not to discard biometrics, but to reclaim them and reshape them into tools that reflect the richness of human diversity, rather than distort it. Only then can biometric technology truly live up to its promise: a future where identification is secure, seamless, and fair for all.

References

1. Davitaia, A. (2025). Machine Learning in Voice Recognition: Enhancing Human-Computer Interaction. *Available at SSRN 5329570*.
2. Davitaia, A. (2025). Advancements in Fingerprint Recognition: Applications and the Role of Machine Learning. *Available at SSRN 5268481*.
3. Davitaia, A. (2025). Optimizing Real-Time Traffic Management Using Java-Based Computational Strategies and Evaluation Models. *Available at SSRN 5228096*.
4. Davitaia, A. (2025). Intelligent Finance: The Evolution and Impact of AI-Driven Advisory Services in FinTech. *Available at SSRN 5285808*.
5. Davitaia, A. (2025). Recursive Techniques for Hierarchical Management in Digital Library Systems. *Available at SSRN 5228100*.
6. Davitaia, A. (2025). Enhancing Library Management with Functional Programming: Dynamic Overdue Fee Calculation Using Lambda Functions. *Available at SSRN 5228094*.
7. Davitaia, A. (2025). Choosing Agile SDLC for a Software Development Project Using React,. NET, and MySQL. *Available at SSRN 5215308*.
8. Davitaia, A. (2025). Fingerprint-Based ATM Access Using Software Delivery Life Cycle. *Available at SSRN 5215323*.
9. Davitaia, A. (2022). The Future of Translation: How AI is Changing the Game. *Available at SSRN 5278221*.
10. Davitaia, A. (2024). From Risk Management to Robo-Advisors: The Impact of AI on the Future of FinTech. *Available at SSRN 5281119*.
11. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 81, 77–91.
12. Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (FRVT) Part 3: Demographic effects* (NISTIR 8280). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

13. Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 429–435. https://doi.org/10.1145/3306618.3314244
14. Jain, A. K., Ross, A., & Nandakumar, K. (2011). *Introduction to biometrics*. Springer. https://doi.org/10.1007/978-0-387-77326-1
15. Klare, B. F., Burge, M. J., Klontz, J. C., Vorder Bruegge, R. W., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, 7(6), 1789–1801. https://doi.org/10.1109/TIFS.2012.2214212
16. Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., ... & West, S. M. (2018). *AI Now Report 2018*. AI Now Institute. https://ainowinstitute.org/AI_Now_2018_Report.pdf
17. Garvie, C., Bedoya, A. M., & Frankle, J. (2016). *The perpetual line-up: Unregulated police face recognition in America*. Georgetown Law, Center on Privacy & Technology. https://www.perpetuallineup.org/
18. Simonite, T. (2019, January 25). When it comes to gorillas, Google Photos remains blind. *WIRED*. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/
19. Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
20. Wilson, D., & Whittaker, M. (2021). *Disability, bias, and AI*. AI Now Institute. https://ainowinstitute.org/disabilitybiasai-2021.pdf
21. Osoba, O. A., & Welser, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html
22. NIST. (2021). *Face recognition vendor test (FRVT) ongoing: Part 6A - Performance of automated face recognition algorithms*. https://pages.nist.gov/frvt/
23. Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics*, 24(5), 1521–1536. https://doi.org/10.1007/s11948-017-9975-2
24. Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2). https://doi.org/10.1177/2053951717743530
25. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. fairmlbook.org. https://fairmlbook.org/
26. Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2979–2989.
27. Jain, A. K., & Maltoni, D. (2021). *Handbook of biometric anti-spoofing: Presentation attack detection*. Springer. https://doi.org/10.1007/978-3-030-87809-3
28. Keyes, O. (2019). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–22. https://doi.org/10.1145/3359246
29. Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., ... & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117
30. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. https://doi.org/10.1145/3458723

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.