# Ethical Data Curation for AI: An Approach based on Feminist Epistemology and Critical Theories of Race

Susan Leavy [1,2] Eugenia Siapera [2] Barry O'Sullivan [1,3]

[1]Insight Centre for Data Analytics [2]University College Dublin [3]University College Cork

## Summary

The potential for bias embedded in data to lead to the perpetuation of social injustice though Artificial Intelligence (AI) necessitates an urgent reform of data curation practices for AI systems, especially those based on machine learning. Without appropriate ethical and regulatory frameworks there is a risk that decades of advances in human rights and civil liberties may be undermined. This paper proposes an approach to data curation for AI, grounded in feminist epistemology and informed by critical theories of race and feminist principles. The objective of this approach is to support critical evaluation of the social dynamics of power embedded in data for AI systems.

## Bias and Discrimination in AI

Artificial Intelligence systems that have been found to discriminate, have invariably disadvantaged those already most marginalised in society. The profound consequences of issues with the design of AI systems and their capacity to perpetuate discrimination have resurfaced decades-old debates concerning social justice and the surveillance of racialized and gendered bodies. Cases of race and gender discrimination due to the way individuals were categorised and treated differently by AI algorithms have been uncovered, demonstrating the capacity of AI to mirror and even exacerbate the discriminatory behaviour that civil rights movements have fought against [2, 6, 7]. Research has uncovered bias and discrimination in predictive policing systems, healthcare decision support systems, facial recognition and language technology [9, 8, 4, 1]. Given the evident threat to social justice and fundamental human rights posed by AI systems learning from data, it is clear that a substantial change is required to methods for AI data curation.

We examine how feminist principles along with critical theories of race may incorporated into a critical framework for ethical data curation for AI that examines data within the context of a particular machine learning approach and along with with the raw data itself, examines approaches to sampling, feature selection and data annotation all of which have the capacity to influence and skew data. This work draws upon the work of D'Ignazio and Klein [3] who outlined feminist principles for data science and Jo and Gebru [5], who proposed a new specialisation in data curation for AI based on archival methods.

## Ethical Framework for Data Curation for AI

The approach to ethical curation of data set out in this paper is grounded in feminist epistemology to enable the interrogation of theoretical standpoints and concepts underlying data for AI systems.

Principles of feminist theory and critical race theories are drawn upon to formulate an ethical framework to mitigate discrimination and bias in data for AI. The following presents a summary of the principles we set out:

### Examine perspectives in data

The perspectives of individuals involved in both data creation and the development of AI systems are embedded in curated data collections and play a central role in the social construction of concepts. Aligning with principles of feminism and critical race theory, to prevent damaging social constructs being learned by AI systems, the first point of critical examination in ethical data curation is to understand who's perspectives are encoded in data and the potential implications of this.

### Recognise the reflexive nature of knowledge

Feminist epistemology recognises that choices made in representing knowledge play a role in generating societal concepts. Aligning with feminist principles and critical race theory, an activist stance in the curation of data for AI is called upon.

### Analyse theory in data

From the standpoint of feminist epistemology, how we represent knowledge in data is heavily value-laden and influenced by philosophical viewpoints. Uncovering the nature of theoretical concepts, particularly pertaining to identity, is therefore central to ethical data curation for AI.

### Include subjugated & new forms of knowledge

Knowledge from groups considered subordinate along with forms of knowledge outside the predominate forms of discourse are often marginalised or omitted from what is considered legitimate forms of data and can therefore be more difficult to access. To address such imbalances, ethical data curation requires inclusion of multiple sources and forms of knowledge.

## Conclusion

The critical framework for the ethical curation of data for AI set out in this paper aims to enable an interrogation of the power structures and theoretical conceptions of identity that may be embedded in data for machine learning rather than addressing specific instances of bias. The assumptions of feminist epistemology upon which this framework is grounded along with principles of critical race theory address how the values and perspectives of those involved in the creation, collection, processing and curation of data for AI are embedded in its contents. The framework emphasises the importance of identifying groups whose knowledge may be omitted and re-balancing data accordingly. The value-laden nature of data and how this can generate bias in AI systems is examined and those involved in the development of AI systems are called upon to adopt an activist stance in the ethical curation of data. Through developing this critical framework for data curation we aim to contribute towards a virtue ethics among technology developers that would facilitate transparency and promote accountability for issues of algorithmic discrimination in AI.

## References

[1] Joy Buolamwini and Timnit Gebru.
Gender shades: Intersectional accuracy disparities in commercial gender classification.
In *Conference on Fairness, Accountability and Transparency*, pages 77--91, 2018.

[2] Amit Datta, Michael Carl Tschantz, and Anupam Datta.
Automated experiments on ad privacy settings.
*Proceedings on privacy enhancing technologies*, 2015(1):92--112, 2015.

[3] Catherine D'Ignazio and Lauren F Klein.
*Data feminism*.
MIT Press, 2020.

[4] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou.
Word embeddings quantify 100 years of gender and ethnic stereotypes.
*Proceedings of the National Academy of Sciences*, 115(16):E3635--E3644, 2018.

[5] Eun Seo Jo and Timnit Gebru.
Lessons from archives: strategies for collecting sociocultural data in machine learning.
In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 306--316, 2020.

[6] Anja Lambrecht and Catherine Tucker.
Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads.
*Management Science*, 2019.

[7] Safiya Umoja Noble.
*Algorithms of oppression: How search engines reinforce racism*.
nyu Press, 2018.

[8] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan.
Dissecting racial bias in an algorithm used to manage the health of populations.
*Science*, 366(6464):447--453, 2019.

[9] Rashida Richardson, Jason M Schultz, and Kate Crawford.
Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice.
*NYUL Rev. Online*, 94:15, 2019.