MANCHESTER 1824
The University of Manchester

AI Ethics and Data

Louise A. Dennis
University of Manchester

MANCHESTER 1824
The University of Manchester

Algorithmic Bias

Algorithmic bias is a socio-technical phenomenon. Its social aspect comprises the biases that have long existed in society affecting certain groups such as underprivileged and marginalised communities, whereas its technical facet involves the manifestation of social biases in algorithms' outcomes. (Kordzadeh & Ghasemaghai 2022)

- Buolamwini & Gebru: **Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.** FAT 2018: 77-91
- Kordzadeh & Ghasemaghaei (2022) Algorithmic bias: review, synthesis, and future research directions, European Journal of Information Systems, 31:3, 388-409, DOI: 10.1080/0960085X.2021.1927212

MANCHESTER 1824
The University of Manchester

How does Bias enter an Algorithm?

- Historical Bias – data reflects historic social bias e.g., in recruitment practice
- Data Selection – some groups may be under-represented in the data set
- Algorithmic Design Bias – choice of cost-benefit optimization may lead to “expensive” groups being ignored
- Human oversight – The human interprets the algorithm’s output through the prism of their own bias.

Centre for Data Ethics and Innovation, Review into Bias in Algorithmic Decision Making, 2020.

MANCHESTER 1824
The University of Manchester

Mitigations

- Datasheets for Datasets: Gebru et al, Datasheets for datasets. Commun. ACM 64(12): 86-92 (2021)
- Explainable AI: de Bruijn et al, The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making, Government Information Quarterly (2022)
- Including fairness metrics in training – but it is challenging even to define fairness. Kleinberg et al, Inherent Trade-Offs in the Fair Determination of Risk Scores. ITCS 2017: 43:1-43:23
- Diverse Teams.

MANCHESTER 1824
The University of Manchester

Assurance: IEEE Standard P7001: Transparency of Autonomous Systems

Winfield et al, IEEE P7001: A Proposed Standard on Transparency. Frontiers in Robotics and AI, section Ethics in Robotics and Artificial Intelligence. 2021.

Validation Transparency (Level 3) includes the need for reports on any "analysis of communities or environments that could be affected by the decisions of the system and the impact on those communities and environments, even where those communities and environments are not explicitly recognized as stakeholders." or a statement that no such analysis took place.

MANCHESTER 1824
The University of Manchester

Other Issues

- Monitoring of users – data-fiction of individuals, surveillance culture, moral/legal reporting requirements.
- Data security – trade-offs with communication speed and bandwidth. Assume the existence of malicious actors.
- Accident Risk Predictions – trade-offs with who/how many are harmed and how much, rules of the road, trolley problems of various kinds.
- Aggregate data – transparency vs. privacy. Zang & Bolot. 2011. Anonymization of location data does not work: a large-scale measurement study. MobiCom '11.