

2025

Algorithmic Impact Assessments

Dr Hadley Beresford, Prof Jo Bates

Information School, University of Sheffield



Contents

- Brief talk on algorithmic bias
- An introduction to algorithmic impact assessments (AIAs)
- 3 activities on algorithmic impact assessments;
 - What type of questions are useful for DWP?
 - How do these questions fit into organisational processes? Who is needed to answer them?
 - Which of these questions are most relevant and prompt reflective thinking?



About the AlgoBias Toolkit project

Aims:

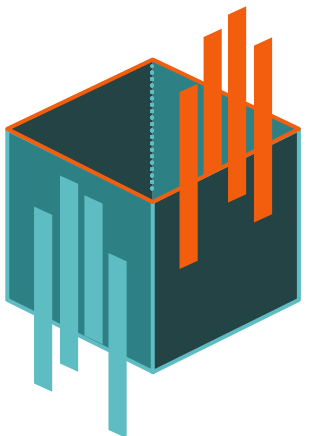
- **Co-design AI toolkit for use at DWP**
- Identify key issues for responsible AI at DWP
- Create a creative commons resource for other civil service orgs and local authorities to use

Project plan

- WP1 Workshops (Feb-March 2025)
- WP3 AIA Toolkit Development (April - June 2025)
- WP4 Pilot & Initial Evaluation of AIA Toolkit (2025)
- WP5 Longitudinal evaluation of workshops and toolkit, and planning next steps (2025)

Content:

- Algorithmic Bias
- Socially Meaningful Transparency
- Algorithmic Impact Assessments



Examples of algorithmic bias

BBC Your account Home News Sport Weather iPlayer Sounds

NEWS

Home | InDepth | Israel-Gaza war | War in Ukraine | Climate | UK | World | Business | Politics | Culture

Family & Education | Young Reporter

A-levels: Anger over 'unfair' results this year

© 13 August 2020



PA MEDIA

This year's exams were cancelled because of the Covid-19 pandemic

By Sean Coughlan, Katherine Sellgren, Judith Burns
BBC News family and education

Artificial intelligence (AI)

Warnings AI tools used by government on UK public are 'racist and biased'

Transparency campaigners welcome government move to publish details of system algorithms

Jon Ungoed-Thomas and Yusra Abdulahi

Sun 25 Aug 2024 06:00 BST

Share

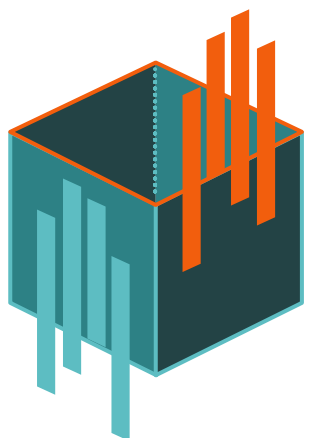


In 2020 a legal challenge stopped the Home Office from using an algorithm to help sort visa applications which contained 'entrenched racism and bias'. Photograph: Guy Corbishley/Alamy

Artificial intelligence and algorithmic tools used by central government are to be published on a public register after warnings they can contain "entrenched" racism and bias.

Officials confirmed this weekend that tools challenged by campaigners over alleged secrecy and a risk of bias will be named shortly. The technology has been used for a range of purposes, from trying to detect sham marriages to rooting out fraud and error in benefit claims.

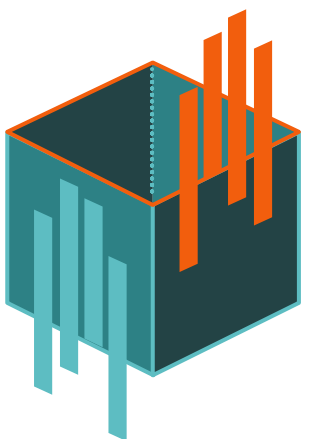
The move is a victory for campaigners who have been challenging the



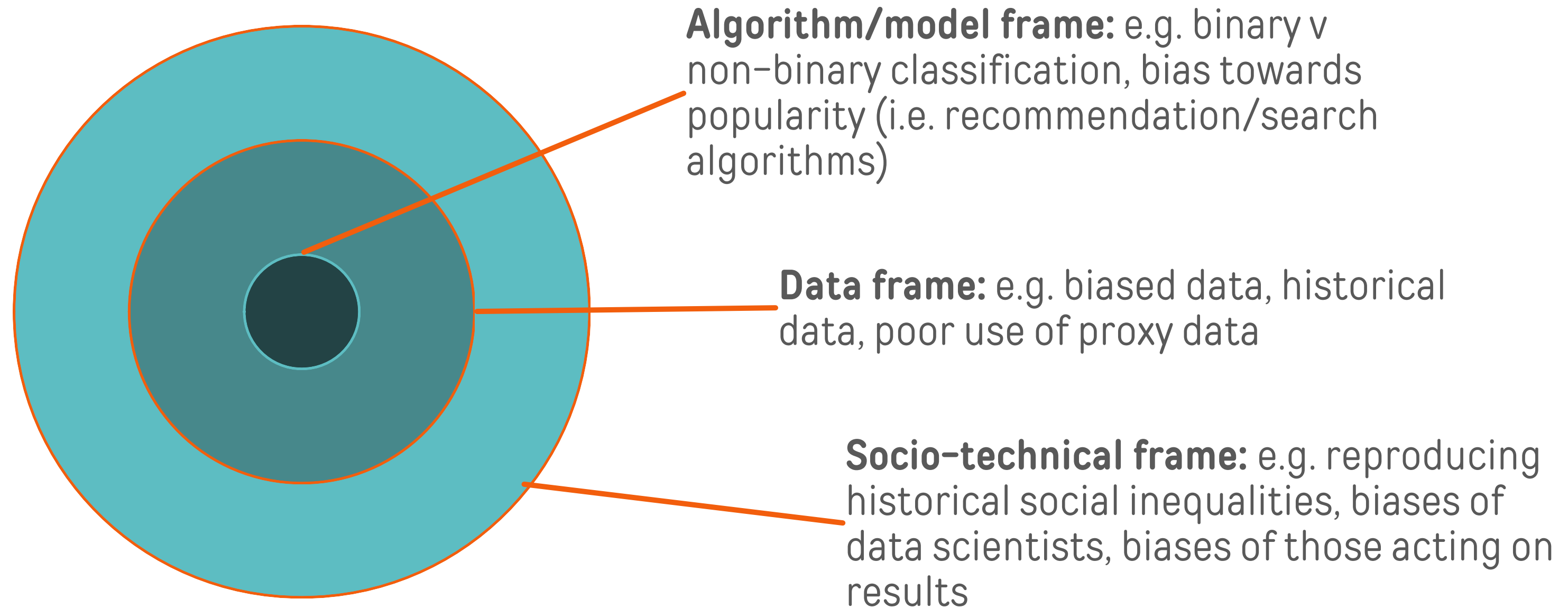
Definition of Algorithmic Bias

- “unfair outcomes caused by the automation of statistical techniques which aim to predict outcomes based on historical data and probabilistic understandings”

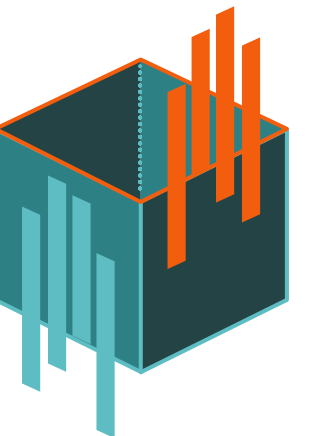
(Beresford, 2020)



Where does algorithmic bias come from?

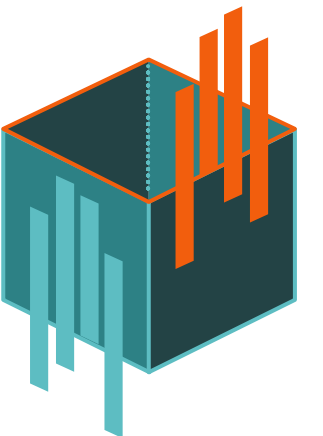


(J. Bates, 2018, adapted from Selbst, Boyd, Friedler, Venkatasubramanian, & Janet, 2018)



Algorithmic bias

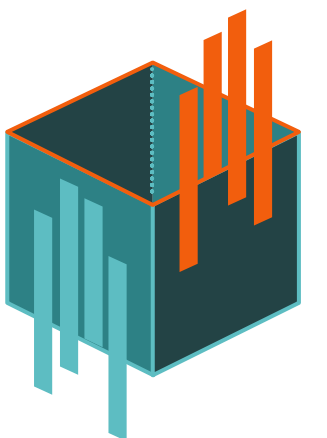
- My PhD research was in partnership with the DWP investigating algorithmic bias (2018-2023)
- Key recommendations for practice are:
 - Practitioners from different disciplines and roles need to **create a shared understanding** of algorithmic bias
 - UK civil servants seeking to further social justice conceptualisations of ‘fairness’ **must scope out potential routes within the constraints of the civil service**
 - Practitioners should **seek to adopt socio-technical algorithmic bias mitigation methods**, such as Algorithmic Impact Assessments(AIAs), Value Sensitive Design (VSD), and critical thinking about data and their wider environment



- Data Protection Impact Assessments (DPIAs) from the GDPR
- Like DPIAs, detailing the risks and how these might be mitigated
- “AIAs are an emerging mechanism, proposed as a method for **building algorithmic accountability**, as they have the potential to help build public trust, **mitigate potential harm and maximise potential benefit of AI systems**” (Ada Lovelace Institute, 2022)
- “AIAs **will not solve all of the problems** that automated decision systems might raise, but they do provide an important mechanism to **inform the public and to engage policymakers** and researchers in productive conversation.” (AI Now Institute, 2018)



About Algorithmic Impact Assessments



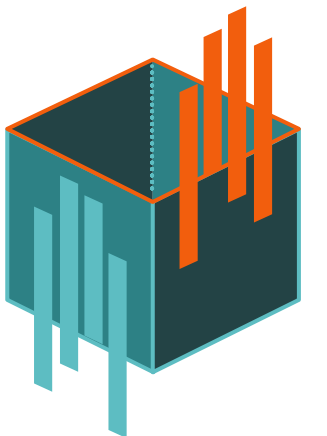
Results from Ada Lovelace and NHS' National Medical Imaging Platform (NMIP) pre-pilot analysis:

- Greater accountability / Transparency
- Reflection / Reflexivity
- Standardisation
- Independent scrutiny / Multi-stakeholder engagement
- Documentation
- Participation / meaningful opportunity to respond or dispute a system

(Adapted from Ada Lovelace 2022)



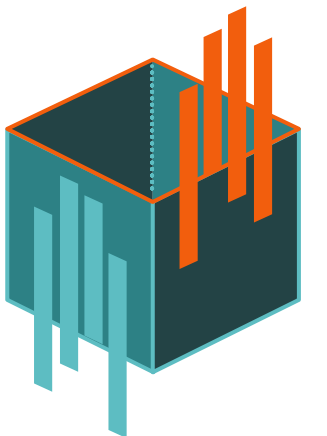
Today



About Algorithmic Impact Assessments



- You'll be presented with a list of questions adapted from the Canadian Open Data Algorithmic Impact Assessment and the Ada Lovelace assessment
- As a team, go through each of the exercises
- For each exercise, full instructions are on the Miro Board
- Password:
- I'm here to facilitate and answer any questions as needed :)



—

Thank
You!

