

A multi-stakeholder perspective on the regulation and design of algorithm fairness

Ansgar Koene, Virginia Portillo & Liz Dowthwaite – University of Nottingham
 Helena Wedd & Menisha Patel – University of Oxford

Summary of stakeholder workshop

UnBias investigates the user experience of algorithm driven services and the processes of algorithm design to provide policy recommendations, ethical guidelines and a ‘fairness toolkit’ that will be co-produced with stakeholders. We focus on a wide range of stakeholders and carry out activities that

- 1) support user understanding about algorithm mediated information environments,
- 2) raise awareness among providers ‘smart’ systems about the concerns and rights of users, and
- 3) generate debate about the ‘fair’ operation of algorithms in modern life.

As part of the project we ran a multi-stakeholder engagement event involving academics, NGOs and SMEs to share perspectives and seek answers to key project questions such as:

- What constitutes a fair algorithm?
- What kinds of (legal and ethical) responsibilities do internet companies have to ensure fair and unbiased algorithmic decisions?
- What factors might enhance users’ awareness of, and trust in, the role of algorithms in their online experience?
- How might concepts of fairness be built into algorithmic design?

The discussion was structured around four case studies:

- 1) gaming the system - anti-Semitic autocomplete and search results;
- 2) news recommendation and fake news;
- 3) personalisation algorithms;
- 4) algorithmic transparency.

Defining algorithm fairness

Candidate definition: “A context-dependent evaluation of the algorithm processes and/or outcomes against socio-cultural values. Typical examples might include evaluating: disparity between best and worst outcomes; sum-total or worst-case outcomes equal treatment without prejudice or advantage due to task-irrelevant factors.”

Criteria relating to:

- Social norms and values
- System reliability
- Non-interference with user control/agency

Regarding design:

- Transparency
- Duty of care to society
- Duty of care to target users/customers

Criticism:

- Algorithms are simply tools, fairness depends on behaviour of actors who use them, their objectives and methods.



General observations

- Bias and unfairness in algorithms is broad in scope with potential to disproportionately affect vulnerable users.
- There is currently no effective regulation or market pluralism of online platforms.
- Algorithms on online platforms can greatly benefit users.
- Effective regulation requires accountability & responsibility from platforms and agencies + meaningful transparency of the algorithms.

Case study 1: Gaming the system

Dec 2016 Carole Cadwalladr wrote in the Observer about anti-Semitic autocomplete and search results for “are Jews” on Google, with a discussion about ranking manipulation through ‘gaming the system’.

Participant suggestions

- There is lack of awareness about how rankings are determined and what they mean.
- Censorship regulation focuses on removal not placement in ranking.
- User input to rate ranking could help signal difference between ranking (by match to search term) vs. validity of content.

How to bump Holocaust deniers off Google’s top spot? Pay Google

Google ‘is unhappy’ with Holocaust denial beating the truth in its search results - but it probably makes more money that way



Prisoners being transported to Auschwitz, 1944. Photograph: Universal History Archive/UG via Getty Images

Case study 2: Fake news

Participant suggestions

- Fake news is not new, there is lack of evidence of impact.
- Solutions should focus on education, critical reading skills, Trustmark/branding & breaking link with financial profit.
- Market research suggestions people don’t want personally tailored news.

Facebook fires trending team, and algorithm without humans goes crazy

Module pushes out false story about Fox’s Megyn Kelly, offensive Ann Coulter headline and a story link about a man masturbating with a McDonald’s sandwich

“I worked on Facebook’s Trending team - the most toxic work experience of my life”



The trending module was meant to have ‘learned’ from the human editors’ curation decisions. Photograph: Bloomberg/Bloomberg via Getty Images

Case Study 3: Personalisation algorithms

Issues

- Creation/amplification of echo chambers
- Inaccurate or discriminatory results
- Profiling and manipulation of users

Participant suggestions

- Personalisation should be called *task based channelling*, it does not consider aspirations.
- Personalisation OK for objects and commercial purposes, not appropriate for socially important information or news
- Users should have control over level of personalisation



Case Study 4: Algorithm transparency

In June 2008 Reddit made their algorithm open source to provide transparency. Most other services are reluctant to do so due to commercial concerns and fear that it makes it easier to game the algorithms.

Participant suggestions

- Meaningful transparency depends on the purpose of transparency (focus on fairness or the trust?).
- Data is integral to determining bias, so transparency also depends on data
- Need to understand users as well as the algorithm and data.
- Auditing and certification by intermediary organisation to give transparency shielded against gaming the algorithms.

