Multi-Stakeholder Dialogue for Policy Recommendations on Algorithmic Fairness

Helena Webb
University of Oxford
Department of Computer Science
Oxford, UK
helen.webb@cs.ox.ac.uk

Ansgar Koene
University of Nottingham
Horizon Digital Economy Research Institute
Nottingham, UK
a.koene@nottingham.ac.uk

Menisha Patel
University of Oxford
Department of Computer Science
Oxford, UK
menisha.patel@cs.ox.ac.uk

Elvira Perez Vallejos
University of Nottingham
NIHR Nottingham
Biomedical Research Centre
Nottingham, UK
e.perez@nottingham.ac.uk

ABSTRACT

Multi-stakeholderism [1] is a valuable methodology for governance and policy development. We describe the use of the approach in the UnBias study, which seeks to identify opportunities for effective governance of algorithmic online services. We use the multi-stakeholder methodology to bring together experts from relevant sectors including academia, education, government, regulation, law, civil society, media, and industry and commerce. This paper outlines how our work has facilitated open and constructive debate that can drive the development of meaningful policy recommendations. We also describe some challenges of this approach and our next steps towards producing actionable design and policy recommendations, including engagement with international industry standards development.

CCS CONCEPTS

• **Social and Professional topics → Computing/technology policy → Commerce policy →** *Consumer products policy*; **Social and Professional topics → Computing/technology policy → Government technology policy→** *Governmental regulations*

KEYWORDS

Multi-stakeholder; Algorithms; Fairness; Deliberation; Social Media; Responsible Research and Innovation

1 ALGORITHMS, MULTI-STAKEHOLDERISM AND POLICY

1.1 Introduction

We live in an age of ubiquitous online data collection, analysis and processing. Social media sites, search engines and recommendation sites increasingly use personalisation and filter algorithms to determine the information we see when browsing online. Whilst these algorithms can help us to cut through the mountains of available information, there are increasing concerns that they can have negative effects – for instance by facilitating the spread of fake news [1] and growth of filter bubbles [2] as well as invading users' privacy [3]. These concerns raise important questions over the appropriate governance and regulation of online platforms and their algorithms. In this work-in-progress paper we describe ongoing work in the UnBias study using a multi-stakeholder approach to explore problems around the use of algorithms on social networks and other online sites, towards identify policy recommendations for their appropriate use. We draw on the emerging findings of our work to demonstrate how soliciting the views of stakeholders from multiple sectors can produce nuanced discussion and constructive debate to drive the development of meaningful policy recommendations.

SMSociety '18, July 18–20, 2018, Copenhagen, Denmark

ACM ISBN 978-1-4503-6334-1/18/07…$15.00
https://doi.org/10.1145/3217804.3217952

1.2 The Multi-Stakeholder Approach for Policy Development

Since the mid-1990s the multi-stakeholder approach has gained increasing popularity as a methodology for governance and policy development. It has underpinned activity in Corporate Social Responsibility (CSR) [4, 5], become a guiding principle at the World Economic Forum [6] and a dominant format for negotiations on internet governance [7] at the OECD, Council of Europe, ITU, and the primary UN forums related to Internet Governance - the IGF and World Summit on the Information Society (WSIS). As summarised by the Internet Society [8], it works best on "messy" (interdependent, complex, emergent) issues where:

- decisions impact a wide and distributed range of people and interests,
- there are overlapping rights and responsibilities across sectors and borders,
- different forms of expertise are needed, such as technical expertise, and
- legitimacy and acceptance of decisions directly impact implementation.

All of the above apply to the problem of producing trustworthy, fair and accountable algorithmic services.

The multi-stakeholder approach is based on the overall notion that those most impacted by a change, issue or circumstance should be involved in the management and governance and ultimately the resolution of that issue. In the case of online algorithmic services, the relevant stakeholders to include in are tech-companies, government regulators, researchers/academics, educators and civil-society groups (representing citizens). Inclusiveness is the basis of legitimacy. The less inclusive a process is, the less likely it is to engender the trust and support of those outside of the process. Transparency is essential for inclusiveness, as it brings experts and affected groups into the process. The most effective decisions are those based on open and deliberative processes that consider a broad range of information sources and perspectives. An important aspect of the process is that all participants have equal opportunity to express their opinions and be heard. Ideally, as in the case of the WSIS forum, the agenda and process formation is also open to be shaped by all the stakeholders.

The multi-stakeholder model for policy development is in effect a governance oriented version of the principles that are at the heart of Responsible Research and Innovation (RRI): *"a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)".* [9]

The model is therefore also being used by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems both for the development of their "Ethically Aligned Design" document [10] and in the associated IEEE P70xx Standards Projects such as the IEEE P7003^{TM} Standard for Algorithmic Bias Considerations that our own project is directly contributing to.

2 UNBIAS STAKEHOLDER ENGAGEMENT

2.1 The UnBias project

The UnBias project seeks to promote fairness online by taking up current concerns about algorithmic processes on popular platforms such as social media sites. We investigate the user experience of algorithm driven Internet services and the processes of algorithm design. We ask key questions such as: are algorithms ever "neutral" and how can we be sure they are operating in our best interests? How can we judge the trustworthiness and fairness of systems that heavily rely on algorithms?

Our project activities adopt a variety of data collection and methodological approaches. A fundamental aim is to engage with perspectives of stakeholders from relevant professional sectors. These include: academia, education, government, regulatory agencies, law, civil society groups, media, and industry and commerce. Importantly, UnBias provides a space where stakeholders can come together to explore implications of algorithm-mediated interactions on online platforms. We follow the multi-stakeholder approach to encourage open discussion of relevant issues in order to harness collective expertise and begin to formulate solutions and develop policy recommendations.

2.2 UnBias Stakeholder Workshops

To ensure the relevance and diversity of the UnBias stakeholder panel we continuously draw on emerging findings from our other workpackages to identify sectors to be included. We use a semi-targeted snowball recruitment strategy that includes contacting professional networks, notifications on mailing-lists, publication of press-releases, and networking at academic conferences, multi-stakeholder forums and public engagement events. The panel currently includes representatives from 55 organisations/academic labs (3 corporate, 5 small-medium enterprises, 5 regulatory bodies, 8 NGOs, 4 schools, 7 consultancies/think tanks/professional associations and 22 academics from Engineering, Computer Science, Business, Education, Social Science and Law).

The primary means through which our panelists engage with the project is by attending stakeholder workshops and online questionnaires. We conducted a series of highly constructive events that helped drive our project findings and received positive feedback from participants. The workshops take different forms but are all designed to encourage wide-ranging discussion of issues connected to the contemporary ubiquity of filtering and personalisation algorithms on online platforms. In this paper we focus on the first two workshops we ran in the study. Workshop 1 focused on fairness in relation to algorithmic practice and design. It was attended by over 30 stakeholder panel members. Participants completed a short pre-questionnaire soliciting their responses to a proposed definition of fairness and their own experiences with using and/or designing algorithm driven systems. The workshop discussed current controversies regarding fake news, personalisation mechanisms, search engines, and algorithmic transparency. Workshop 2 was attended by 20 panel members and began with a limited-resource-allocation task asking participants to consider dilemmas of algorithmic fairness in relation to a specific scenario. We then had them to discuss their views in relation to a planned project output: an empathy tool to help algorithm designers to better understand the user perspective.

During both workshops we paid careful attention to building an open and inclusive atmosphere that encouraged wide ranging debate and constructive disagreement. We also applied the Chatham House Rule, which states that comments made in the workshops can be repeated but only without being ascribed to the individuals who made them. As a result, both events generated fruitful discussions that highlighted the existence of complexities and multiple perspectives in relation to the design, development and use of algorithms. With the consent of participants, and University Research Ethics Committee clearance, the workshops were audio recorded and transcribed. We also provided post-workshop questionnaires for feedback.

3 PRELIMINARY RESULTS AND DISCUSSION

3.1 Overview

The outcomes of the first two stakeholder workshops provide an opportunity to begin assessing the value of a multi-stakeholder approach to discussing controversies around algorithmic online services and its potential to generate policy recommendations. We describe four key features of our workshops that particularly demonstrate the benefits of this approach. These features, illustrated with anonymised examples from both workshops, were identified by analysing completed pre-questionnaires, workshop transcripts, feedback questionnaires and our own observations and reflections. We then describe certain practical and conceptual challenges associated with this approach.

3.2 Findings

3.2.1 Constructing the messy problem. In their (written and spoken) contributions participants consistently constructed the question of algorithmic regulation, and the component issues that constitute it, as a “messy problem”. They referred to the question as complex and hard to resolve, calling on the existence of multiple perspectives and stances. To give one example, our pre-Workshop 1 questionnaire included a working definition of algorithmic fairness: “*...a context-dependent evaluation of the algorithm processes and/or outcomes against socio-cultural values. Typical values might include evaluating: the disparity between best and worst outcomes; the sum-total of outcomes; worst-case scenarios; everyone is treated/processed equally without prejudice or advantage due to task-irrelevant factors.”*

We invited participants to rate and comment on this definition. Although most rated it “good” or “a reasonable starting point” they offered several suggestions for improvement. For instance, they made suggestions regarding system reliability – such as the need to balance results with due regard for trustworthiness – and social norms and values – such as balancing of individual values against collective ones. Participants also suggested the need to recognise the importance of user agency and freedom from interference, for instance in terms of users being able to limit the data that is collected about them or to opt out of an algorithmic process that is not relevant to the tasks they want to perform. With these responses participants collectively highlighted the nuance and complexity surrounding the concept of fairness when applied to algorithm-driven online platforms. Fairness was constructed as a messy problem and by extension so too was the problem of effective regulation. The ways in which participants constructed the problem in their discussions was useful to us as researchers as it made visible the range of, sometimes conflicting, issues that are central to questions of regulation. Similarly, it was helpful to our participants as it enabled them to access alternative perspectives and fostered genuine debate.

3.2.2 The combination of perspectives from multiple sectors. Participation of people from a broad range of sectors was fundamental to our outcomes since individuals frequently contributed comments reflecting their own expertise. The combination of perspectives from multiple sectors enabled detailed discussion that attended to both the nature of the messy problem and its potential solutions from a range of relevant angles.

For instance, Workshop 1 included a discussion of “fake news” in relation to the ways in which algorithmic processes on social media platforms can facilitate the rapid spread of unverified content. Participants with a legal background focused on the (lack of) regulation of social media platforms, in comparison to traditional news media as, a component part of the problem. By contrast stakeholders from platforms and commerce tended to focus on the user experience. During the discussion of potential solutions, participants from various backgrounds highlighted the value of educating users and suggested that critical thinking could be taught in schools as a particular mechanism to counter fake news. Whilst this idea received a lot of support amongst the group, a teacher present sounded a note of caution: "*the idea of educating people about fake news may be so problematic politically speaking that it can't happen. Because ... we have to demonstrate as teachers that we are politically neutral in the manner in which we deliver our education. ... So if we are going to start discussing whether or not a piece of news is fake or not, very often we will get into political territory which we are really not supposed to be engaging in, [W]ithin the education system it would be very complicated to start 'educating' people about fake news. We can give them the tools and the skills to detect bias and so forth, but we can't start necessarily, talking about trust in sources without getting into sticky waters to do with politics."*

As this example shows, the provision of contributions from various areas of stakeholder expertise builds up a nuanced discussion of the issues at hand. It also facilitates correction where assumptions are made by stakeholders from one sector about what is or is not possible in another. In this way, discussions are inclusive and pay respect to different relevant sectors.

3.2.3 Combining the abstract and the particular. Across both workshops we asked stakeholders to discuss concepts including fairness, justice and empathy as well as specific case studies of current controversies regarding algorithmic online services. This provided an excellent means to combine the abstract and the particular when we analyse different viewpoints relating to algorithmic processes and policy.

We notice that stakeholders frequently move between the abstract and the particular when expressing their views, with one sometimes used to reinforce a point made about another. For instance, comments on our working definition of fairness, referenced specific systems which could be seen as unreliable or not providing adequately for user control. Similarly, discussion of algorithmic transparency moved between i) what transparency means in an abstract sense and how it could help to better understand how and why certain content is being shown, and ii) specific instances in which transparency could bring benefits – e.g., fostering greater trust in news feeds on social media – or disadvantages – e.g., enabling some users to 'game' search algorithms etc. to their own ends.

A frequently expressed frustration was that abstract concepts of fairness etc. do not adequately allow for the specific circumstances in which problems occur. As one participant commented in relation to a discussion of potential bias in algorithm design: "*Well if you want to discuss about fairness, you have to go into a causal analysis to try and work out your context and what actually is important. You cannot have a one size fits all rule*."

By expressing this frustration over what is lost in the move from the particular to the abstract our stakeholders highlighted an important challenge regarding the development of policy and policy recommendations. Policy is by necessity general rather than specific but to what extent is it ever possible to find a satisfactory and generic, one-size-fits all solution?

3.2.4 The value of dialogue. Both workshops generated genuine dialogue in which participants exchanged viewpoints and countered arguments by putting forwards alternative perspectives and evidence. This enabled the discussion to delve deep into the nuance and complexities of the issues at hand. This dialogue sometimes came about through participants making points tied to their own areas of expertise. In other instances, dialogue was not sector specific and instead participants shared and developed viewpoints based on personal understandings and perceptions. In Workshop 2 participants were given a questionnaire task that required them to select a preferred algorithm for the distribution of limited resources in a specific scenario. Participants completed the task individually and then discussed their answers as a group. The sharing of perspectives on what constituted fairness in the given scenario led some participants to change their preference selections and also reconsider whether fairness could be best conceptualised as equality of opportunity or equity of outcomes, requiring prioritisation of some..

In the same workshop, the dialogue between participants helped us refine our own project design. We asked our participants to discuss our idea for an empathy tool, a material artefact for online providers and other stakeholders to help them understand the concerns and rights of internet users. We asked participants to consider what form this tool might take. The discussion that ensued highlighted issues we had not thought of. We had assumed that empathising with users would lead developers to make positive, supportive changes to algorithms and/or platforms. However, our stakeholders pointed to alternative evidence - such as Facebook's advertising campaigns to exploit users' emotional states [11] – demonstrating that empathy does not necessarily lead to positive action and can increase manipulation. The profit models of many platforms can in fact be seen to incentivise this manipulation. One participant described:

"⋯*the mismatch of empathy to internet firms. I mean when their business model is to extract every single thing they possibly can from us, for their own financial gain and benefit, how likely is it, or even reasonable that we can encourage them to empathise with their users, who essentially are just a piggy-bank to raid?"*

As a result of this dialogue we decided to refine our project design and change our plans for an empathy tool. This demonstrates the deep value of dialogue in fostering productive exchanges that can underpin deliberation and meaningful improvements.

3.3 Practical and Conceptual Challenges

The outcomes of our workshops highlight the value of the multi-stakeholder approach. These workshops generated an exchange of cross sector perspectives revealing the “messiness” of problems arising from the ubiquity of algorithmic online services. They also fostered genuine dialogue and consideration of relevant matters both in the abstract and the particular. This inclusive approach paves the way to identify points of consensus and to develop policy recommendations that legitimately represent a broad stakeholder perspective. However, the multi-stakeholder approach does present some obstacles and challenges to be overcome.

There are a number of practical challenges. Most professionals are very busy so it can be very difficult to find a time and place that suits everyone. The use of online questionnaires and remote participation can help, but we have still been unable to solicit input from all members of our stakeholder panel. Face-to-face events are undoubtedly the most productive as they enable real time discussion and dialogue. However, they are expensive and labour intensive as they require multiple facilitators to work with subgroups, lead and annotate discussions etc. Furthermore, some organisations can be wary of allowing their members to join events where they their sector may be criticised by others present. Disagreement amongst stakeholder participants is inevitable - and in fact is to be encouraged as it can create a pathway that leads to constructive ideas for change. However, it needs to be handled delicately so that all participants feel their views are welcome and respected.

At the conceptual level one key challenge for the approach is the problem of retaining scope limits. The richness of perspectives elicited during the process can easily produce valid arguments for expanding the scope of policy regulation beyond the specific goals of the session and initiative; for instance, our workshop discussions about monitoring for unintended algorithmic outcomes (relating to the spread of unverified content or development of filter bubbles etc.) often spread into discussions about socio-economic inequalities. Facilitation of multi-stakeholder dialogue therefore requires a fine balance between being receptive to unexpected perspectives (as benefitted us in discussion of the empathy tool) and maintaining focus, or at least direction, on the target topic. A further conceptual challenge occurs when consensus or agreement across participants is not possible. In particular, if different sectors are in disagreement, is there any way to move forwards without appearing to give one sector greater validity than another? This is a highly complex issue and resolving it is often contingent on the aims and format of the specific initiative and policy topic being considered.

4 CONCLUSIONS

Our work-in-progress illustrates how open and inclusive multi-stakeholder discussion can provide a rich source of insight into messy problems, such as the regulation of algorithmic online services. This multi-stakeholder approach is well-suited to a Responsible Research and Innovation based project design, and generates avenues of academic inquiry that are embedded in real-world relevance. As the UnBias project continues we will distil the findings of these, and future, workshops into policy recommendations and actionable steps for inclusion in design recommendations and the IEEE P7003 Standard for Algorithmic Bias Considerations. Our recommendations will particularly benefit from the nuanced perspectives put forward by our stakeholder panel members.

ACKNOWLEDGMENTS

We thank all members of the UnBias project stakeholder panel. UnBias is funded by the Engineering and Physical Sciences Research Council (ref: EP/N02785X/1).

REFERENCES

[1] Sapna Maheshwari. 2016. How fake news goes viral. A case study. *The New York Times* online at https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html

[2] Sally Adee. 2016. Burst the filter bubble. *New Scientist*, 232, 24-25.

[3] Alex Hern. 2018. Breach leaves Facebook users wondering: how safe is my data? *The Guardian* online, 18 March

[4] Jerry M. Calton and Steven L. Payne. 2003. Coping With Paradox: Multistakeholder Learning Dialogue as a Pluralist Sensemaking Process for Addressing Messy Problems. *Business and Society* 42 (1), 7-42. DOI: 10.1177/0007650302250505.

[5] Luc W. Fransen and Ans Kolk. 2007. Global rule-setting for business. A critical analysis of multi-stakeholder standards. *Organization* 14(5), 667-684. DOI: 10.1177/1350508407080305

[6] World Economic Forum. 2010. *Annual Report 2009-2010* http://www3.weforum.org/docs/WEF_AnnualReport_2009-10.pdf

[7] World Summit on the Information Society. 2003. *The multi-stakeholder participation in WSIS and its written and unwritten rules*. https://www.itu.int/net/wsis/basic/multistakeholder.html

[8] Internet Society. 2003. Internet Governance: *Why the multistakeholder approach works*.

[9] René von Schomberg (Ed.). 2011. *Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields*. Luxembourg: Publication Office of the European Union.

[10] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2017. *Ethically Aligned Design - Version II*

[11] Sam Machkovech. 2017. Facebook helped advertisers target teens who feel "worthless" *Ars Technica. 5/1/2018.*