# Algorithmic Accountability, Trustworthiness and the Need to Develop new Frameworks

*Farida Vis, Research Fellow in the Information School at the University of Sheffield, investigates the issue of trust in the debate about algorithmic accountability, arguing that we should instead focus on 'trustworthiness' and that now is the time for a considered debate about algorithmic governance and accountability frameworks.*

For 2016 the Oxford English Dictionary word of the year may very well turn out to be 'algorithm'. I have, along with many others, noticed how this word had started to seep into everyday language more and more, but this year feels like a turning point (see for example this recent article in Slate). Think of the furore over the introduction of an algorithmic timeline on Twitter, which will no longer be organised chronologically, but rather by 'importance' of content. Somehow it felt that this concept now needed less explanation than it might have done a year ago. From looking at some of the comments on #RIPTwitter, where users discussed the rumour before the confirmation a few days later, the news was dominantly interpreted along these lines: Twitter would become more like Facebook, which in the opinion of this vocal group, should be avoided. There are also concerns for research associated with this recent development, but in the context of algorithmic accountability, the way in which Facebook content is shown in user's feed is a longstanding example of why there is a need for more accountability on social media platforms. This issue will serve as the focus of this post.

As a company, Facebook is not well trusted, something highlighted in a recent write up of Prophet's new 'Brand Relevance Index'. The index, recently reviewed by Adweek, captured just how lowly users rate Facebook in terms of trust. 'Asked to rate Facebook as "a brand I can depend on," respondents ranked it at 133. And as "a brand I can trust," Facebook fell lower still, to 200'. But when it comes to writing about algorithmic accountability, academics and advocates often cite trust as a key pillar of focus: one should strive for more trust, implying that 'more trust is good, less trust is bad.' There is a similar focus on making things more transparent and open. It is worth considering critically what a focus on these issues in particular means for how we might think about algorithmic accountability and how this might push us to develop critiques in a particular, in more limited way.

In terms of thinking about trust, I was recently reminded of work by Onora O'Neill, who highlights best what is at stake. O'Neill suggests we need more trustworthiness rather than trust. This distinction forces us to consider normative dimensions—precisely what characterises trustworthiness—rather than focus on measuring whether more or less trust exists. (O'Neill's 2002 Reith Lectures 'A Question of Trust' give a good overview).

If we briefly return to Facebook's low rating in the Brand Relevance Index, this serves as a great example of O'Neill's distinction and the need to move beyond an emphasis on trust. What is clear is that users don't trust Facebook, but they still use it. You could say: yes, but that's because Facebook has a monopoly, users have no choice of comparable social networking sites. Fair enough. The point though is that users can often say one thing: 'I don't trust', but make daily decisions whereby they still use the products of companies they claim not to trust.

In terms of thinking about algorithmic accountability, I think it's therefore important that we question the frameworks that are used and mobilised in this context. Are the accountability frameworks that often get adopted 'fit for purpose'? Or are accountability frameworks pushing us

down well-trodden terrain? I would argue that there is a need for a more considered debate about these issues.

As part of my work with Pia Mancini on the World Economic Forum's Global Agenda Council on Social Media, we have started to think through what different possible frameworks for algorithmic accountability and governance might look like. A key aim is to offer an accessible overview of the breadth of current debates, across various sectors and academic disciplines. In line with Kate Crawford's recent writing, we problematise the seemingly narrow focus on transparency issues, including the often associated methods for investigating or indeed 'opening up' algorithmic black boxes. These methods are frequently premised on the idea that the ways in which the algorithms we encounter on a daily basis behave (through our social media newsfeeds for example) are knowable through basic experiments. Is it certainly the case that we may identify interesting ways in which these algorithms *appear* to behave in relation to 'us' (the individual user), but it would be difficult to do such work more systematically across large groups of different users, thus getting beyond the level of what is essentially anecdotal evidence.

Taking a different approach, our WEF research has a strong focus on creating a bridging function between policy stakeholders and potentially less well-informed end users of technologies such as Facebook. We explore what a range of different governance scenarios might entail, who they might include and how they can serve a wider range of affected stakeholders. Therefore we are specifically interested in how we can develop techniques that allow for contributions to these discussions from as broad a range as possible of interested members of the public.

Our initial framework includes twelve different scenarios for algorithmic accountability and governance that we consider important as part of this exploratory work (longer version here). These scenarios range from exploring what's at stake in a 'business as usual' approach, where companies do nothing with respect to their algorithmically driven platforms, to one that would offer full algorithmic disclosure, which we essentially view as an impossibility, given that this is based on the idea that the ways these complex systems work is fully knowable. We outline what different national bodies governing algorithms might look like versus a supra, multi-stakeholder independent body (modelled for example on the World Wide Web Consortium (WC3)). We explore what an algorithmic 'smart contract' between a user and company could entail and how flexible this needs to be, or what might be included in a 'public agreement of values' and a 'dashboard' scenario, where users can opt in/out of various features in order to protect and control their data.

We also explore the different ways in which the creation of value in these systems works: situations where social media users essentially create potentially valuable information (which economists would term 'positive externalities'), and how users can own it and profit from it. Paul Mason, writing about the Spotify data controversy after the introduction of new terms of service last summer, summarises this as 'a third thing produced for free, accidentally, by the interaction of buyers and sellers'. We consider such examples of positive externalities to explore where an 'algorithms as a public utility' scenario might possibly take us, and reconsider how and by whom value may be extracted, and under what conditions.

In the context of the earlier points about 'trustworthiness', our sixth scenario is worth quoting in full: 'This scenario sets out a set of agreed values, generated following public discussion about what is acceptable behaviour on the part of platforms, including how data is algorithmically sorted and presented to users. This public agreement of values thus cover what is considered ethical, fair as well as legal.' The next phase of this work will concentrate on how to best facilitate these discussions, so that as many interested people can take part in them, and not only already well-informed researchers.

*This article gives the views of the author, and does not represent the position of the LSE Media Policy Project blog, nor of the London School of Economics.*

*This post was published to coincide with a workshop held in January 2016 by the Media Policy Project, 'Algorithmic Power and Accountability in Black Box Platforms'. This was the second of a*

*series of workshops organised throughout 2015 and 2016 by the Media Policy Project as part of a grant from the LSE's Higher Education Innovation Fund (HEIF5). To read a summary of the workshop, please click here.*

March 7th, 2016 | Algorithmic Accountability, Featured | 0 Comments