Selected Papers of #AoIR2018:
The 19th Annual Conference of the
Association of Internet Researchers
Montréal, Canada / 10-13 October 2018

TOWARDS FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN PLATFORM GOVERNANCE

Robert Gorwa
University of Oxford

Background

Algorithms, especially those of the machine learning variety, are playing an increasingly significant role in modern life. Billions of people interact every day with complex algorithmic systems like Google's search engine (PageRank) or Facebook's news feed (EdgeRank), and algorithmic decision-making is now being deployed in high-stakes domains such as policing, finance, and healthcare (Ananny & Crawford, 2016). Galvanized by the constant flow of "algorithmic war stories" that illustrate how bias and discrimination can be exhibited by these systems (Edwards & Veale, 2017), a growing community of social and computer scientists has been working to establish actionable frameworks for accountable algorithms and fair machine learning, often summarized as Fairness, Accountability, and Transparency in Machine Learning, or FAT-ML (Barocas & Selbst, 2016).

A similar line of research has yet to be pursued in the emerging area of platform governance, a body of work that addresses technology "platforms" as socio-technical systems and political actors that engage in private governance (Gillespie, 2017; Denardis & Hackl, 2015). Through their design decisions, terms of service, and content policies, platforms intervene in social and political activity (Gillespie, 2015). These policies not only influence what users see and do, but they also can have broader social and political implications: most recently, Facebook, Google, and Twitter have been accused of playing a problematic role in the lead up to the 2016 US Election, with critics claiming that their policies enabled the malicious use of targeted advertising, propaganda, and automated accounts ("bots").

Given that, as Caplan and boyd (2018: 2) observe, "conversations around algorithmic accountability often center on Facebook and Google," there is an unexplored nexus between the algorithmic accountability literature and the platform governance literature. Focusing on platforms – not just how they are governed, but also how they themselves govern (Gillespie, 2017) – offers insight into several important questions for FAT-ML scholars. How can fairness, accountability, and transparency be truly achieved when

Suggested Citation (APA): Gorwa, R. Towards Fairness, Accountability, And Transparency In Platform Governance. (2018, October 10-13) Paper presented at AoIR 2018: The 19th Annual Conference of the Association of Internet Researchers. Montréal, Canada: AoIR. Retrieved from http://spir.aoir.org.

the overarching structures within which these algorithms are embedded may not be fair, accountable, or transparent? Likewise, the algorithmic accountability literature has much to offer those interested in contemporary technology platforms. For instance: what do we now demand from algorithms, that we do not similarly demand from the corporate entities that deploy them?

A FAT Framework for Platforms

Combining insights from political theory, political philosophy, and international relations with the burgeoning algorithmic accountability and fair machine learning literatures, this paper will seek to present a novel perspective on platform governance. I set out to answer the following question: to what extent is fairness, accountability, and transparency a useful framework for thinking about platform governance?

Focusing on Facebook as a case study, and drawing on a variety of qualitative data collected in 2018 – participant observation in a multi-day "policy deep dive for researchers" organized by Facebook executives, as well as interviews conducted with current and former Facebook policy employees – this paper will critically analyze Facebook's policy practices within a fairness, accountability, and transparency framework. I also explore how Facebook employees themselves discursively conceptualize the notion of fairness, accountability, and transparency in their day-to-day work.

Preliminary Findings & Arguments

I generally find that FAT provides an imperfect but useful heuristic for thinking about questions relating to social media policy, while also serving as an important normative goal as to what platform governance should strive to achieve.

Thinking explicitly about fairness, transparency, and accountability can yield interesting insights into how the policy processes of platforms really function in practice. For example, Facebook employees appear to exhibit a deep desire that their policies be considered "fair," in the narrow context of non-discrimination across individual users. In content moderation, Facebook strives for reproducibility and consistency in moderation, meaning that the same content posted by two users would be equally likely to be either pulled down or left online moderators, despite the difficulty of achieving this in practice. However, Facebook policy teams have more difficulty with a broader notion of fairness, one that involves balancing individual rights to expression against possible societal, macro-level harms spread across a population. (Should Facebook, in the context of hate speech, consider "white" as equivalent to other minority groups, given the historic patterns of discrimination within the United States? See Angwin and Grassegger, 2017).

Investigating accountability suggests that Facebook policy processes are responsive to the public in certain ways (especially public backlash and "PR fires," which could be said to constitute a narrow form of "public accountability"), but ultimately are limited in their accountability to the public due to their corporate structure and incentives. The paper explores Facebook's recent (and ongoing) efforts to become more transparent, while also showing how Facebook policy is a kind of "black-box," with the effect that process of policy decision-making, and the motives and goals behind these policies are not transparent or clear to the users who are directly affected by them.

The modern social media ecosystem involves the interplay of a multitude of varying (and often competing) public and private interests, making platform governance both incredibly complex and important. Significant issues are at stake: as Denardis and Hackl (2015: 9) argue, the core functions of social media platforms are "simultaneously related to several conditions of democracy: how people receive news; the articulation of relationships and associations; access to knowledge; and spaces for deliberation about issues of public concern."

The questions raised in this paper – effectively, how these decisions are made, and whether it is possible to make them in a less problematic manner – are important ones for the internet research community. The inner workings of platforms are still not well understood, and as Annany and Crawford (2016: 10) explain, "A system needs to be understood to be governed." This paper hopes to provide an empirical jumping off point for what promises to be a major area of research, debate, and discussion in the years to come.

References

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 1461444816676645.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Cal. L. Rev.*, *104*, 671.

Binns, R. (2017). Fairness in Machine Learning: Lessons from Political Philosophy. *Journal of Machine Learning Research*, *81* (2018 Conference on Fairness, Accountability, and Transparency), 1–11.

Caplan, R., & boyd, danah. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society*, *5*(1), 2053951718757253.

DeNardis, L., & Hackl, A. M. (2015). Internet governance by social media platforms. *Telecommunications Policy*, *39*(9), 761–770.

Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review*, *16*(1), 18–84.

Gillespie, T. (2010). The Politics of "Platforms." *New Media & Society*, *12*(3), 347–364.

Gillespie, T. (2015). Platforms Intervene. *Social Media + Society*, *1*(1), 2056305115580479.

Gillespie, T. (2017). Governance of and by platforms. In J. Burgess, A. Marwick & T. Poell, eds., *The SAGE Handbook of Social Media.* Los Angeles, CA: SAGE.

Angwin, J. & Grassegger, H. (2017, June 28). Facebook's Secret Censorship Rules Protect White Men From Hate Speech, But Not Black Children. *ProPublica*. Retrieved from https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms