



Robin Burke

Abstract The incorporation of fairness-aware machine learning presents a challenge for creators of personalized systems, such as recommender systems found in e-commerce, social media, and elsewhere. These systems are designed and promulgated as providing services tailored to each individual user's unique needs. However, fairness may require that other objectives, possibly in conflict with personalization, also be satisfied. The theoretical framework of post-userism, which broadens the focus of design in HCI settings beyond the individual end user, provides an avenue for this integration. However, in adopting this approach, developers will need to offer new, more complex narratives of what personalized systems do and whose needs they serve.

1 Introduction

The turn toward questions of fairness in machine learning (Barocas and Selbst 2016; Dwork et al. 2012; Mitchell et al. 2021) raises some important issues for the understanding of personalized systems. Researchers studying these systems and organizations deploying them present a common narrative highlighting the benefits of personalization for the end users for whose experience such systems are optimized. This narrative in turn shapes users' expectations and their *folk theories* (working understandings) about the functionality and affordances of personalized systems (DeVito et al. 2017). Fairness requires a different kind of analysis. Rather than focusing on the individual, fairness is understood in terms of distribution: how is harm or benefit from a system distributed over different individuals and/or different classes of individuals? These distributional concerns take the focus at least partly away from the end user, and thus the implementation of fairness concerns in personalized systems requires a re-thinking of fundamental questions about what personalized systems are for and what claims should be made about them.

R. Burke (✉)
University of Colorado, Boulder, CO, USA
e-mail: robin.burke@colorado.edu

A personalized system tailors its presentation to an evolving understanding of what a user seems to want or need in the present moment. What are we to do when the thing that the user wants contributes to unfairness? To make this question more concrete, let us consider a particular class of personalized system, recommender systems, which filter and prioritize information and items for users in a personalized manner. Common examples include applications embedded in social media, streaming audio and video, news, job hunting, and e-commerce sites. Consider a recommender system embedded in an employment-oriented site such as XING or LinkedIn that supports human resource professionals in locating job candidates for positions. It may be that the user (the recruiter in this case) by virtue of their interactions builds a profile that disadvantages applicants in a protected category: women, for example. The recruiter's recommendation lists are disproportionately filled with male candidates, and they might not even realize the sequence of events bringing this about. Worse yet, the prevalence of male candidates will likely cause the recruiter's interactions to be primarily with male applicants, generating even more evidence (from the system's perspective) of their interest in these candidates over female or non-binary ones. The system thereby becomes what O'Neil (2016) calls a "weapon of math destruction," a computationally governed feedback cycle generating ever societally worse outcomes.

From the standpoint of pure personalization, this outcome might appear to be a success story: the user tells the system what they want and the system delivers, getting "better" at doing so over time. However, it should be clear that this is a highly undesirable outcome. A system that perpetuates systemic unfairness, even if only by passing on biases from its input to its output, by definition becomes itself part of the oppressive system (Kendi 2019). If we are interested in the beneficence of personalization, we cannot ignore this risk, and therefore we are called up on to re-consider the concept of personalization itself. The concept of "post-userism" as articulated by Baumer and Brubaker (2017) is a theoretical approach that calls into question the user focus that has dominated study in human-computer interaction since the field's inception and raises the possibility that, in understanding and evaluating computing systems, a larger and more complex framing may be essential. Our somewhat dystopian, although hardly unrealistic, personalized recruiting system suggests precisely that we need to look beyond the end user to understand how to build recommender systems free from such harmful effects.

2 De-centering the User

To de-center the user in recommendation is to consider the possibility of additional stakeholders whose objectives and goals should be integrated into the generation of recommendations. The concept of the stakeholder emerges in the management literature in the mid-1960s (as a contrast to the shareholder) defined by some authors as "any groups or individuals that can affect, or are affected by, the firm's objectives" (Freeman 2010). A multiplicity of non-user considerations, especially business

objectives, have entered into practical recommender systems designs from their first deployments. However, businesses that employ such recommendation objectives have generally been very reluctant to identify anything other than user benefit as a driver for their technical decisions. An explicit consideration of multistakeholder objectives and one that specifically incorporates fairness is much more recent (Abdollahpouri et al. 2020).

What might a post-userist take on personalization look like? We examine multistakeholder considerations that can help answer this question.

First, consider the recruiter-oriented recommender system outlined above. The challenge here is to achieve *provider-side fairness* (R. Burke 2017): fair representation across those providing items to be recommended, in this case the job applicants. So, a desired system design is one in which there are limits to the degree of personalization that can be performed, even for a single user.

The system would need to ensure that each recommendation list has at least a minimal degree of diversity across different protected group categories. One would expect that an end user/recruiter would need to know going in (and might even require as a matter of law or company policy) that the recommender is enforcing such fairness constraints.

A more relaxed version of this provider-side constraint might appear in a consumer taste domain, such as the recommendation of music tracks in a streaming music service. The organization might have the goal of fair exposure of artists across different demographic groups or across different popularity categories (Mehrotra et al. 2018). List-wise guarantees might not be important, as there may be some users with very narrow musical tastes and others who are more ecumenical. As long as the goal of equalizing exposure is met, the precise distribution of that exposure over the user population might be unimportant. In this case, it may be desirable to differentiate between types of users for the purposes of fair exposure as in Liu et al. (2019) or to more precisely target the types of diversity of interest to individual users (Sonboli et al. 2020). A system that worked in this manner might need to inform users that non-personalization objectives such as fairness are operative in the recommendations it produces.

An important distinction between the cases above is that music tracks are non-rivalrous goods: a music track can be played for any number of users, and its utility is unaffected by the number of recommendations or their occurrence in time. A job candidate is different. A highly qualified candidate may be present in the job market for a very limited period of time. A recruiter who is recommended such a candidate as soon as their resume appears in the system gets greater utility from the recommendation than does a user who gets it later. A situation in which a highly qualified candidate appears only to a limited number of recruiters is more valuable to them than a situation in which their recommendations are shared with a larger group. One could imagine that the recruiter-users would be rightly concerned that the inclusion of multistakeholder considerations in recommendation could put them at a disadvantage relative to the purely personalized status quo. I say “might” here because the close study of multistakeholder recommendation is sufficiently new that it is unclear what the interactions are between recommendation quality as

experienced by users and the fairness properties of the associated results. Preliminary results in some applications indicate that simultaneous improvements in both dimensions may be possible (Mehrotra et al. 2018). Where there is a tradeoff between accurate results and fair outcomes, we may need to consider the distribution of accuracy loss as a fairness concern across the users of the system (Patro et al. 2020).

The picture changes when we consider *consumer-side fairness*. Here we are interested in fairness considerations across the end users themselves, and this requires a community orientation in how the recommendation task is understood. We can draw on employment as an example yet again, now in terms of recommending jobs to users.

The tension between personalization and other goals becomes complex when we consider that users' behaviors themselves, the raw data over which personalization operates, may themselves be subject to measurement inconsistency. For example, female users are known to be less likely to click on ads that contain masculinized language about job performance (e.g., "rock star programmer"), but this says nothing about their underlying capabilities for such jobs (Hentschel et al. 2014). Even more fundamentally, there may be differences among users in their experience of the data gathering required by personalized systems; individuals experiencing disempowerment may identify the surveillance capacities that enable personalization as yet another avenue of unwanted external control (V. I. Burke and R. D. Burke 2019).

Even if we postulate that profiles can be collected in a fair and acceptable manner, it still may be the case that a system performs better for some classes of users than others. Improving fairness for disadvantaged groups may involve lower performance for others, especially in a rivalrous context like employment, where, as noted above, recommending something to everyone is not desirable. For example, a recommender system might optimize for fairness in such a way that a particularly desirable job is shown more often to members of a disadvantaged group and less often to others. How should a user think about bearing some of the burden in terms of lower utility of providing fairer recommendations to other users in the system? Accepting such behavior in a system requires adopting a pro-social orientation toward the community of fellow platform users, something that may not be easy to cultivate in a multistakeholder recommendation context.

Finally, we should note that the perspectives of recommendation consumers and item providers do not exhaust the set of stakeholders impacted by a recommender system. Authors such as Pariser (2011) and Sunstein (2018) have noted the way in which algorithmic curation of news and information has potential far-reaching impacts on society and politics. Incorporating this wide set of stakeholders draws the focus of a recommender system even further from personalization as a defining characteristic.

3 Conclusion

This discussion shows that the need to incorporate fairness into personalization systems requires more than just technical intervention. The way users think about these systems will need to change radically, and the onus lies on technologists to provide new terminology and new narratives that support this change. A key step may be to acknowledge the multistakeholder nature of existing commercial applications in which recommendation and personalization are embedded and to challenge the simplistic user-centered narratives promulgated by platform operators.

Acknowledgments This work was supported by the National Science Foundation under Grant No. 1911025.

References

- Abdollahpouri, Himan et al. (2020). “Multistakeholder recommendation: Survey and research directions”. In: *User Modeling and User-Adapted Interaction* 30, pp. 127–158.
- Barocas, Solon and Andrew D Selbst (2016). “Big Data’s Disparate Impact”. In: *California law review* 104.3, p. 671-732.
- Baumer, Eric PS and Jed R Brubaker (2017). “Post-userism”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6291–6303.
- Burke, Robin (July 2017). “Multisided Fairness for Recommendation”. In: Workshop on Fairness, Accountability and Transparency in Machine Learning. arXiv: 1707.00093 [cs.CY].
- Burke, Victoria I and Robin D Burke (2019). “Powerlessness and Personalization”. In: *International Journal of Applied Philosophy* 33.2, pp. 319–343.
- DeVito, Michael A, Darren Gergle, and Jeremy Birnholtz (2017). ““Algorithms ruin everything” #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media”. In: *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 3163–3174.
- Dwork, Cynthia et al. (2012). “Fairness Through Awareness”. In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ITCS '12. New York, NY, USA: ACM, pp. 214–226.
- Freeman, R Edward (2010). *Strategic management: A stakeholder approach*. Cambridge University Press.
- Hentschel, Tanja et al. (Jan. 2014). “Wording of Advertisements Influences Women’s Intention to Apply for Career Opportunities”. In: *Academy of Management Proceedings* 2014.1, p. 15994.
- Kendi, Ibram X (2019). *How to be an antiracist*. One World.
- Liu, Weiwen et al. (2019). “Personalized Fairness-aware Re-ranking for Microlending”. In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19. ACM.
- Mehrotra, Rishabh et al. (2018). “Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems”. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 2243–2251.
- Mitchell, S., Potash, E., Barocas, S., D’Amour, A. and Lum, K., 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, 8, pp. 141-163.
- O’Neil, Cathy (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Patro, Gourab K et al. (2020). “Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms”. In: *Proceedings of The Web Conference 2020*, pp. 1194–1204.
- Sonboli, Nasim et al. (July 2020). “Opportunistic Multi-aspect Fairness through Personalized Re-ranking”. In: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '20. Genoa, Italy: Association for Computing Machinery, pp. 239–247.
- Sunstein, C.R. (2018). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

