Summer 2022

The Multisided Complexity of Fairness in Recommender Systems

Nasim Sonboli
University of Colorado, Boulder

Robin Burke
University of Colorado, Boulder

Michael Ekstrand
Boise State University

Rishabh Mehrotra
Spotify, Inc.

Received: 13 April 2021 | Accepted: 8 July 2021

DOI: 10.1002/aaai.12054

SPECIAL TOPIC ARTICLE

The multisided complexity of fairness in recommender systems

Nasim Sonboli[1] | **Robin Burke**[1] | **Michael Ekstrand**[2] | **Rishabh Mehrotra**[3]

[1]University of Colorado, Boulder, Boulder, Colorado, USA

[2]Boise State University, Boise, Idaho, USA

[3]Spotify, Inc., London, UK

Correspondence
Robin Burke, University of Colorado Boulder, Boulder, CO, USA.
Email: robin.burke@colorado.edu

Present address
Nasim Sonboli, Tufts University, Medford, Massachusetts, USA. Rishabh Mehrotra, ShareChat, London, UK.

Funding information
National Science Foundation, Grant/Award Number: IIS-1911025

Abstract

Recommender systems are poised at the interface between stakeholders: for example, job applicants and employers in the case of recommendations of employment listings, or artists and listeners in the case of music recommendation. In such multisided platforms, recommender systems play a key role in enabling discovery of products and information at large scales. However, as they have become more and more pervasive in society, the equitable distribution of their benefits and harms have been increasingly under scrutiny, as is the case with machine learning generally. While recommender systems can exhibit many of the biases encountered in other machine learning settings, the intersection of personalization and multisidedness makes the question of fairness in recommender systems manifest itself quite differently. In this article, we discuss recent work in the area of multisided fairness in recommendation, starting with a brief introduction to core ideas in algorithmic fairness and multistakeholder recommendation. We describe techniques for measuring fairness and algorithmic approaches for enhancing fairness in recommendation outputs. We also discuss feedback and popularity effects that can lead to unfair recommendation outcomes. Finally, we introduce several promising directions for future research in this area.

INTRODUCTION

Recommender systems learn from users' interests and behaviors in order to infer their preferences and provide them with recommendations tailored to their interests. Well-known examples are associated with social media apps (Twitter, Facebook), streaming media services (Spotify, YouTube), and e-commerce sites (Amazon, Etsy), among others. A variety of machine learning techniques have been employed to implement such systems including nearest-neighbor methods (Ning, Desrosiers, and Karypis 2015), sparse matrix and tensor factorization (Aggarwal 2016; Ricci, de Gemmis, and Semeraro 2012), probabilistic reasoning (Barbieri and Manco 2011; Kouki et al. 2015), and most recently deep learning (Zhang et al. 2019). Recommender systems are largely characterized by their personalized nature: they deliver results tailored to individual users' needs and preferences (Ricci, Rokach, and Shapira 2011). The emphasis is often on evaluations that capture some measure of user satisfaction. However, in many recommendation domains, besides the receivers of recommendations (consumers, customers, etc.), other stakeholders such as item providers (vendors, content creators, etc.) and the system itself have an interest in the system's performance and behavior. The interests, goals, and needs of all the stakeholders of interest may need to be considered in the recommendation process and the evaluation of its outcomes (Abdollahpouri et al. 2020).

This broader view of recommender system impact coincides with recent developments in assessing the effects of AI generally. As AI systems pervade society, it has become clear that we must consider their potential negative social impacts. Depending on the type of application, these can take a wide variety of forms, some of which have been detailed in recent scholarship (Benjamin 2019; O'Neil 2016). One important category of analysis that has emerged is *unfairness*, broadly defined as harmful disparity in experiences with a system (Barocas and Selbst 2016; Bozdag 2013; Dwork et al. 2012; Pedreshi, Ruggieri, and Turini 2008). Many systems are trained on large corpuses of data collected from human activity, and can incorporate the biases and discrimination found in wider society. Further, data features, learning models, and objective functions are all selected and/or designed by humans, with their own biases and priorities informed by their social and economic context. Recommender systems share all these concerns and are often applied in settings with significant social impact.

However, there are key differences between the problem of fairness as studied in other AI applications and its counterpart in recommendation (Ekstrand et al. To Appear, §4). These differences derive largely from two specific characteristics of recommender systems: *personalization*, the objective of providing customized results to individual users, and *multisidedness*, the property that recommendation interactions often involve multiple types of parties and potentially multiple fairness concerns that may need to be balanced. Because recommender systems are personalized systems, a key element of their functionality is that they reflect individual user preferences in the recommendations that they provide. Those preferences, or the observations of them by the system, may in themselves contain biases, and therefore fair outcomes might be in tension with the personalization objective. Because recommender systems are often embedded in multisided platforms (Evans and Schmalensee 2016), their stakeholders can include both individuals receiving recommendations and individuals whose items are being recommended. Fairness concerns may, therefore, arise for stakeholders on each side and these may need to be balanced.

These properties are not unique to recommendation or information access settings; they can arise in other AI applications as well. For example, speech recognition and personalized medicine customize their inferences based on the user or patient. Therefore, the intersection of personalization and fairness is likely to be of concern for AI developers beyond the recommendation scenarios that we discuss here. In addition, it is worth noting that major deployed AI-intensive applications (think Siri or Google Translate) are not software applications that users can "own" in any sense. These are cloud applications under constant monitoring and development by their respective owners and, especially since they are offered at no cost to end users, it is reasonable to expect that these applications serve a variety of stakeholders and balance multiple objectives across them (Zuboff 2019). Thus, the complexity of fairness that arises in multisided recommendation settings may also be found in these and similar large-scale AI systems.

Even without the additional complexities of personalization and multisidedness, fairness itself is a complex and, some have argued, an *essentially contested* concept, for which we cannot expect a simple or stable definition (Mulligan et al. 2019). The pursuit of fairness in any sociotechnical system must be responsive to the complexity and multivocal nature of this concept and the potential pitfalls involved in social computing (Selbst et al. 2019). Thus, we emphasize a dynamic, iterative approach to fairness and one that allows for the bottom-up emergence of *fairness concerns*, specific ways a system may provide an outcome or treatment that violates fairness-related objectives.

Figure 1 shows a sketch of the processes involved in developing and operating a fairness-aware recommender system. The process begins with an organization and possibly external stakeholders engaging in consultation to uncover fairness concerns. Those concerns become formalized as metrics, measurable properties of the system that quantify its fairness relative to each concern. The metrics are important to both algorithm creation and system evaluation. On the system side, they provide guidance about what interventions should be employed; on the evaluation side, they indicate whether fairness objectives have been met, which is important feedback to the stakeholders from whom the fairness concerns arise.

In practice, adopting the above process entails understanding and quantifying nuanced trade-offs between different stakeholder metrics, and gauging the impact of algorithmic interventions on stakeholders associated with a particular fairness concern as well as other stakeholders via direct or indirect effects. The challenge of applying these concepts in practice is that the full scope of stakeholder impacts may be only apparent over the long term. For example, the *calibration* method of Steck (2018) tries to generate recommendations whose content characteristics (distribution of musical genres, for example) match those of the user's profile, so users get recommendations whose breadth matches their expressed interests. While this objective is focused on user outcomes, enhancing calibration in a system where it is low will likely have benefits for artists in niche genres, who might otherwise

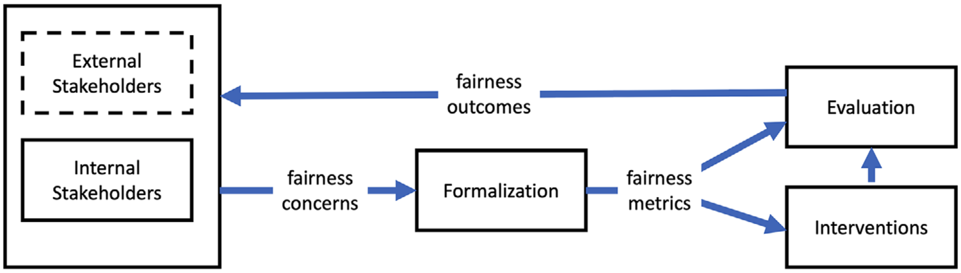

FIGURE 1 Processes in developing fairness-aware recommendation

find themselves rarely heard when recommendations are focused on more popular parts of the catalog.

There is a great deal of complexity in applying ideas from algorithmic fairness in recommender systems. For simplicity, and to reflect the state of current research, we will make two assumptions about fairness concerns. First, we will assume that an organization developing a recommender system will be primarily concerned with two types of stakeholders: *consumers* of recommendations, and *providers* of items or information being recommended (Burke 2017). Fairness concerns may arise for stakeholders of either type (and sometimes for both at the same time). This first assumption concentrates our attention on *distributional fairness*: the distribution of the system's benefits across different individuals and stakeholder groups.

Distributional fairness can be contrasted to *representational fairness*, which looks at how a system may unfairly impact data subjects. For example, a recommender system for commercial imagery may show only white, male models in its photo recommendations (Karako and Manggala 2018), thus presenting a distorted, homogeneous, view of its subject matter and reinforcing the marginalization of those omitted. Here the fairness concern is not a function of the immediate impact of the system on users looking for photos or on the photographers who created them, but rather the harm created by reproducing the minoritization of nondominant groups or viewpoints (Noble 2018). Many of the techniques described here associated with provider fairness may also be applicable to representations fairness problems, but such problems remain understudied in the research literature.

The second assumption is that, when a fairness concern is focused on subgroups of consumers or providers, the grouping can be characterized as binary, separating individuals into a *protected* group for whom fairness is sought relative to an *unprotected* group. This is a great oversimplification of the complex and textured nature of fairness concerns in many real-world settings, but it is typical of machine learning fairness research to date. This simplification is inspired by U.S. antidiscrimination law, when there is a historically discriminated-against group or groups that the law is seeking to protect.

MULTISIDED PLATFORMS

Recommender systems are important components in platforms that bring together buyers and sellers, creators and audiences, and other groups. Economists recognize multisided platforms as systems where search, interaction, and transaction costs are reduced when multiple types of actors join together, thereby providing value (Evans and Schmalensee 2016; Rochet and Tirole 2003). Multisided platforms encompass a wide range of businesses and value propositions and it is, therefore, difficult to generalize about them. Such platforms will employ recommender systems in order to reduce search costs for one (or possibly multiple) sides of the interaction, providing more convenient access to the platform's resources.

Many multisided platforms have only two key types of actors (buyers and sellers, for example) but examples exist in which additional actors are needed to complete the platform's transactions. For example, digital furniture platform Opendesk connects buyers, designers, and producers on its platform. Consumers (buyers) looking for quality craft furniture go to the website where a crowd-sourced community of artisans (designers) have uploaded designs that local manufacturers (producers) can create. Similarly, the food delivery marketplace UberEats consists of eaters, restaurant-partners, and delivery-partners, as the three sides of its marketplace. Fairness concerns may arise for any of the stakeholder groups in these scenarios and, while we do not examine such examples in depth, the issues discussed here are broadly applicable to these more complex platforms.

Here we introduce three examples of multisided platforms and their associated fairness considerations. We will refer to these examples throughout the discussion that follows.

Job recommendation

One prominent example of a multisided recommendation environment is *job recommendation*, as in professional social networks such as LinkedIn (Borisyuk, Zhang, and Kenthapadi 2017); job seekers need to find positions to which they can apply, and want to be considered for good jobs, while recruiters want to find and hire effective and qualified workers. Further, both jobs and candidates are in limited supply: each job can only be filled by a small number of candidates, possibly only one, and each candidate can only work for a small number of employers. As Borisyuk et al. note, naïve recommendations that do not account for these properties may recommend the same jobs to too many candidates, hindering both candidates' ability to find a job opening that is not over-applied, and employers' ability to have prospective candidates discover their posting.

Job and candidate recommendations also complicates somewhat the distinction between consumers and providers, as it can vary based on perspective and on the particular application under consideration: when recommending job listings, job seekers are consumers and employers or recruiters are providers; in a candidate search platform, prospective employees are providers while recruiters are the consumers (Geyik, Ambler, and Kenthapadi 2019). Each of these systems is a separate recommendation problem where the fairness concerns relative to the job seekers and to the employers will manifest themselves in different ways.

Philanthropic recommendation

As e-commerce has become a dominant shopping avenue, so have online sites for philanthropic activity. For example, Kiva is a nonprofit organization that operates a crowd-sourced microlending platform with the goal of financial global inclusion. In Kiva, the sides of the interaction are the borrowers, generally developing-world entrepreneurs who seek small amounts of capital to enhance their business capacity, and lenders, who are the application's end-users. Ab typical lender will contribute only a fraction of the total amount for any one loan, but may support multiple loans at any one time. Lenders do not get any interest on their investments and so supporting a Kiva borrower is essentially a philanthropic act. Kiva's mission emphasizes equitable access to capital for its borrowers, who generally cannot make use of traditional forms of banking and lending (Choo et al. 2014). Lenders are the users of the recommender system, which has the purpose of lowering their search costs in finding borrowers whose goals and needs appeal to them.

Several provider-side fairness concerns might arise in this recommendation context.[1] One key concern, arising from Kiva's mission of supporting world-wide access to capital, is that the geographic imbalances in users' preferences may manifest themselves in the disproportionate representation of certain countries or regions in recommendation lists. This could give rise to a positive feedback loop, as the recommended items are more likely to be supported, and thus the lending becomes even more highly concentrated. A similar kind of imbalance may arise with respect to different industries or economic sectors. Thus, we can identify at least two fairness concerns within this recommender system: equity in geographic distribution of capital (Liu et al. 2019), and equity across economic sectors (Sonboli et al. 2020).

Streaming media recommendation

Streaming services for digital media consumption offer another example of multisided platforms, providing users with on-demand access to global content, thereby connecting users with content creators, for eample, listeners and artists in music streaming platforms such as Spotify, and users and movie makers on video streaming platforms like Netflix. Algorithmically generated recommendations power and shape the bulk of consumption patterns on such platforms and therefore streaming platforms carefully consider the influence of their recommendations on consumption patterns in a manner benefiting not only the users, but also creators, and the long-term goals of the platform itself. As Mehrotra et al. (2018) note, to maximize user satisfaction, streaming platforms optimizing for relevance inadvertently provide impedance to exposure of artists on the tail-end of the spectrum.

To better navigate such trade-offs between different stakeholder objectives, platforms are increasingly relying on multi-objective methods to jointly optimize multiple user-centric goals (e.g., engagement metrics like clicks, number of songs played, time spent), artist-centric goals (e.g., exposure), and platform-centric goals (e.g., discovery, diversity) (Mehrotra, Xue, and Lalmas 2020). In certain cases, aptly balancing such varied objectives makes it possible to obtain gains in both complementary and competing objectives.

ALGORITHMIC FAIRNESS

The extensive recent literature on fairness in machine learning has generally focused on problems of classification: ensuring that when an algorithm makes a decision about the distribution of a benefit (e.g., approval of a loan) or the imposition of a burden (e.g., pretrial detention vs. release on bond) that this decision is fair, in the light of known biases in the historical data on which it is trained. Friedler, Scheidegger, and Venkatasubramanian (2021) make a key distinction between two types of assumptions that a fairness-aware system might make. A system can assume that its observations of features do not encode systematic bias and therefore the goal is to ensure the individuals who are similar along all the relevant decision dimensions are treated similarly. This assumption is described as *What you see is what you get (WYSIWYG)* and can be used to support an approach of individual fairness, where the concern is around the equal treatment of individuals.

The major alternative assumption is one that acknowledges the biases in any data that might be observed and therefore starts from a standpoint that differences between groups, especially groups defined by sensitive features, are a result of societal biases. Friedler, Scheidegger, and Venkatasubramanian describe this as the *We're all equal (WAE)* assumption, which can be used to support a group fairness approach.[2]

Mitchell et al. (2021) provide a thorough treatment of the concepts and metrics—along with vital underlying assumptions—developed to date for measuring and providing fairness in general machine learning, typically classifiers. However, recommender systems have crucial differences in their problem structure that require fairness concepts to be adapted (Ekstrand et al. To Appear, §4); one important one is that recommender systems produce ranked outputs, so decisions about items are not independent (classifier fairness constructs typically assume independent outcomes and decisions). Items are also subject to repeated decisions, unlike most previous fairness problem settings. The personalized and multistakeholder nature of recommendation introduces further complications. Applying WYSIWYG and WAE distinctions, therefore, requires subtlety, and both may be in play in different aspects of the same system or fairness analysis.

WYSIWYG may be regarded as an inappropriate assumption in many recommendation contexts, as there are many well-known biases that skew collaborative signals such as clicks and ratings. However, the fundamental premise of personalized recommendation is that different users have different tastes, and should receive recommendations tailored to those tastes; not every item or provider, therefore, will be relevant to all users. On the other hand, to assume "we're all equal," requires that we clearly define what "all" and "equal" mean in a particular context. Recommender system developers and researchers need to be clear, as Friedler, Scheidegger, and Venkatasubramanian (2021) advocate, about what they are assuming, and how these assumptions follow from an understanding of the underlying fairness issues and concerns.

Fairness to consumers

As we have seen, a recommender system delivers benefits to users by offering them personalized results in the form of recommendation lists, ranked streams, or other outputs. Following Burke (2017), we refer to these individuals as recommendation consumers and fairness concerns involving them will be described as *consumer-side* concerns.

A key consumer-side fairness concern is simply the difference in system performance (for example, prediction accuracy) across different users, especially across groups of users sharing some protected characteristic. Any system will have a range of performance characteristics and may exhibit more errors for some users than others. If these errors are unfairly distributed, especially relative to a protected group, this may be a fairness concern for the system. Yao and Huang (2017) discuss different kinds of error-based unfairness metrics for collaborative filtering. Note that this type of fairness employs the WAE logic: trying to ensure that protected and unprotected users are treated equally by the algorithm.

Rivalrousness or *subtractibility* (Becker and Ostrom 1995) introduces an important subtlety for understanding such distributional fairness concerns. For consumers, some resources are subtractible while others are not; overall utility is generally not subtractible, as one user receiving effective recommendations does not interfere with the relevance or utility of another user's. A case of subtractibility arises in job recommendation, where only a limited number of job opportunities can be recommended to qualified candidates. To avoid some positions receiving too many applications and others too few, the system must balance its promotional efforts in response to the dynamics of the market (Kenthapadi, Le, and Venkataraman 2017). Fair distribution of subtractible recommendations, therefore, becomes a consumer-side fairness concern: since not every compatible job will be shown to every potential job seeker, are members of a protected group getting their fair share of the most desirable recommendations?

Fairness to providers

On the provider side, the utility of a recommender system is that it presents the provider's associated items or information to potential users, buyers, or audiences. We can think of the generation of recommendations for a particular individual as constituting a (subtractible) recommendation opportunity, in which a given provider may or may not participate. Organizations may be concerned that recommendation opportunities are fairly distributed across individual providers or to protected groups of providers. As with the consumer side, recommendation opportunities may be differentially valuable in some absolute sense (purchasers with greater disposable income, for example). A given recommendation opportunity may also be relatively more valuable for some providers than others, by virtue of the characteristics of the associated user.

Examples of fairness concerns

We can see these different types of concerns play out in the examples introduced above. Job recommendation and candidate search bring recommendation explicitly into regulated spaces, at least in many jurisdictions. U.S. employment antidiscrimination law requires that job applicants be treated fairly, without discrimination based on race, ethnicity, gender, or other protected characteristics. This applies at all stages of the job recruiting pipeline; the liability aspects of candidate recommendation are not yet well-established, but fairness towards job candidates is necessary to implement the spirit, if not the letter, of employment law. This means that the side of the platform where the job seeker resides will generally be the focus of any fairness concern in this kind of system.

In the case of Kiva, we noted two examples of provider-side concerns: geography and economic sectors. From historical data, we can determine if there are observable biases among users in what loans are funded quickly on the site, and set as the fairness objective the goal of increasing the recommendation opportunities available to loans in protected regions and sectors.

In the case of music streaming, we can quantify differences in artist exposure for user's organic streaming versus programmed streaming. These differences, especially if associated with protected groups among the artists, could constitute a provider-side fairness concern.

Multisided fairness

Multisided fairness arises when there are fairness concerns on both sides of the recommendation interaction. In such cases, the system must balance such concerns relative to each other. Although the constraints and problem formulation are quite different, this balancing act inherits some characteristics of the classic stable matching problem (Roth and Sotomayor 1992). While a recommender system is collecting and processing user preferences, these profiles are generally incomplete and therefore user preferences are inferred and uncertain. It is also the case that a recommender is delivering more than a single option to each recommendation consumer and the provider typically can satisfy multiple user matches at once (and sometimes an unbounded number).

Specific application and data characteristics must inform system design and optimization where a balance between competing fairness objectives needs to be struck. The close and continuing involvement of stakeholders may be necessary to keep such a system on track. An inspirational example in the algorithmic fairness literature is the WeBuildAI framework, created to ensure participatory stakeholder involvement in algorithm design (Lee et al. 2019).

Multiple fairness concerns create the need for such balancing acts, but similar considerations may arise even if the primary fairness concerns are only on one side of the interaction. It is possible that addressing fairness concerns on one side may change the performance of the system with potential fairness implications on the other. For example, improving fairness for providers may lower the accuracy of recommendations that end-users receive, a kind of "fairness tax." A system may wish to ensure that this disutility is distributed fairly across consumers (Patro et al. 2020). The imposition of this kind of cost (even if fairly distributed) may be challenging in practical contexts due to the fact that users might not share the system's fairness objectives (Modani et al. 2017).

Finally, as noted above, our focus on consumer-side and provider-side fairness does not cover the space of all potential stakeholders. In the UberEats three-sided platform, consumers are users who order food, providers are restaurants who provide the food, but there are also the drivers who make the deliveries. Drivers constitute a *side stakeholder* group, who are impacted by the recommendations the system makes, even though they do not directly participate (Abdollahpouri 2020). We might want to ensure that customers from different parts of the city have similar ETAs for their food delivery (consumer fairness). At the same time, we might want to be fair to restaurants by giving them equal visibility in recommendations and be fair to drivers by distributing the delivery orders in an equitable way.

METRICS AND EVALUATION

The complexity of fairness concepts as outlined above hints at the challenge involved in measuring the fairness properties of a recommender system. It is essential to have the metrics used be aligned with the fairness concerns and benefits or harms that they point towards. Metrics have a primary role of telling system operators and others whether the system is in fact fair in the ways intended, and they also may be useful as optimization objectives or constraints.

Provider fairness

There are a variety of options for measuring provider-side fairness, tied to different kinds of assumptions about the fairness concern itself and the types of benefits that the recommender system has for providers.

One set of metrics starts with the WAE assumption and assumes that provider utility is a function of the appearance of associated items in recommendation lists. Under this rubric, the goal of fairness is to equalize the probability of appearance of protected and unprotected group items in aggregate across recommendation lists. Most simply, we can measure the average proportion of protected group items in such lists. This approach can be extended to take ranking into account, for example by looking at prefixes of the list of increasing length (Yang and Stoyanovich 2017; Zehlike et al. 2017), or by modeling user attention as a function of rank to quantify the utility accruing to protected and unprotected group items (Sapiezynski et al. 2019).

Under a WYSIWYG assumption, we can treat user ratings and, by extension, predictions of the same by the recommender system as indicative of the value of a match between user and item. With this approach, we can couple group exposure in rankings with the relevance of that group's items to the users. This can be done by comparing exposure to relevance, for example, by requiring that a group's exposure be proportional to its relevance (Biega, Gummadi, and Weikum 2018; Singh and Joachims 2018), or it can be done by using relevance data to derive an ideal *target exposure* for items or groups, and comparing the system's actual exposure to that target (Diaz et al. 2020). An ideal system, under this scheme, puts the most relevant items at the top of the ranking; but among comparably relevant items, the exposure is fairly distributed with respect to items' individual providers or their providers' group membership.

Neither of these approaches takes provider preferences into account, the idea that providers themselves might attach different utilities to different recommendation opportunities. Problems of this type have been explored in the area of computational advertising, where advertisers specify directly or indirectly (through bidding), which users are most preferred targets for their messages (Wang and Yuan 2015).

In the example applications discussed above, a variety of fairness approaches and metrics have been defined. In studies of fairness in the Kiva microlending domain, Liu et al. (2019) define a protected category of loans with a lower likelihood of recommendation and calculate the average exposure of these protected group items in recommendation lists. In Mehrotra et al. (2018), fairness is defined relative to artists at different deciles in the popularity distribution and measured over the exposure counts for each popularity group. In job recommendation, there is little published research that makes use of demographic data on job seekers due to the sensitivity of such features. Some researchers have explored fairness as the balance between the preferences of job seekers and employers, employing equilibrium concepts from economics (Xia et al. 2019).

Consumer fairness

Consumer fairness is concerned with fair and equitable treatment of all the users in the system regardless of their membership to any protected group. In this area, researchers have relied exclusively on the WAE assumption; the idea being that users are entitled to similar quality of service from the recommender system. In this family of solution approaches, researchers measure the quality of recommendations that users experience and compare across groups. The differences in these approaches comes from the metric they use to measure system performance, for example, error- or accuracy-based metrics or even diversity-based metrics such as KL-Divergence as in Steck (2018).

For example, Yao and Huang (2017) compare the discrepancy in the recommendation lists of consumer groups, for example, in the average predicted rating of different user groups, or different types of errors among user groups such as overestimation and underestimation of their predicted ratings. Ekstrand et al. (2018) compare the distribution of the authors' genders in user rating profiles and recommendation lists produced from this data and demonstrate their inconsistencies. Ekstrand and colleagues also performed an off-line top-N evaluation of several collaborative filtering algorithms and compared the results for different user demographics (Ekstrand et al. 2018).

These group-wise performance measures do not consider any inherent value associated with recommended items, although this may be a fairness concern in some contexts, as we have noted: higher-salary jobs versus lower-salary ones, higher-risk loans versus lower-risk ones, and so forth. Consumer-side fairness metrics that incorporate the differential utility of recommended items have not yet been explored in the research literature.

Evaluation

Recommender systems can be evaluated using either online or offline methods. For practical reasons, academic researchers have generally made use of offline methods as these require only access to historical data about user preferences, as opposed to online research, which requires access to a large and active user base. However, offline evaluation has significant disadvantages for studying fairness-aware recommendation in particular as the aspects of the system where fairness is important are likely to be those where less data about user preferences will be available. Any evaluation about users' receptiveness to these items will be inherently more uncertain than that of more popular items. In Mehrotra et al. (2018), the authors present a method for gathering historical *exploration data* from a live recommender system, which can then be used to evaluate recommendations across the full range of potential user–item combinations in an unbiased way. This still requires access to a live system, but only once, for the data gathering and not for every algorithm evaluation.

Offline evaluation methodologies entail producing training and test splits of historical user profile data, training the recommender on the training data and evaluating against the test data. A typical methodology is to split each user profile so that there is both training and test data for each user. Since utility is evaluated relative to the known test data items, it can be considered a lower bound for actual utility as experienced by a user: there will likely be recommended items that have utility for a user but were not rated by them.

In an online system, users are presented with recommendations and their reactions are measured with respect to those recommendations. Classic A/B testing techniques are typically employed to determine the properties of algorithms relative to a variety of performance metrics, including fairness. For example, in Mehrotra et al. (2018), user satisfaction is measured as the number of recommended tracks that a user listens to. One of the challenges of employing A/B testing is understanding its impact on user experience. When a fairness objective is in tension with other performance measures, business objectives may require that the fairness aspects of the system constitute only a minor intervention to preserve user experience.

Another valuable type of evaluation is a human subjects study that examines the perceptions users, content creators, and other stakeholders have of recommendation and its fairness aspects. There are relatively few studies of this type for fair machine learning generally; some examples include (Dodge et al. 2019; Harrison et al. 2020; Kasinidou et al. 2021; Srivastava, Heidari, and Krause 2019). There has been even less attention to the recommender system specific issues discussed here. Sonboli et al. (2021) report on such a study, focusing the need for explanations that justify the application of fairness and other non-accuracy objectives and that are grounded in users' pre-existing understanding of recommender system functionality. Ferraro, Serra, and Bauer (2021) report on interviews with music artists to gain their (provider-side) perspective on fair music recommendation.

FAIRNESS-ENHANCING INTERVENTIONS

With fairness concerns identified, it becomes possible to consider interventions to improve recommender systems performance relative to them. In doing so, it is worth keeping in mind the "traps" identified in Selbst et al. (2019), possible hazards in applying fairness concepts in sociotechnical systems. Sebst and colleagues note that sometimes a technical fix is not always the most appropriate approach for problems of power imbalance and bias, and the failure to recognize this is defined as the *solutionism trap*. There may be a wide variety of non-computational solutions to problems that surface themselves as unfair recommendations.

Preprocessing

Preprocessing methods focus on compensating for the existing biases in a dataset. Chen, Johansson, and Sontag (2018) suggest different data collection enhancements to compensate for the biases that occur due to data imbalance. For example, if there is an under-represented group in the data, more data can be collected or imputation can be used to give the system a more complete picture. Resampling data so that there is balanced representation of groups can be useful for testing whether data imbalance is a contributor to metric imbalance; doing so in MovieLens data has been seen to reduce the discrepancy in recommendation accuracy for users of different genders (Ekstrand et al. 2018). However, if resampling increases the overall sparsity of a data set, it may induce a trade-off between fairness and accuracy similar to other types of interventions.

In-processing

In-processing approaches try to improve the fairness of results by integrating fairness notions into recommendation generation itself. Approaches can include multi-objective optimization, constrained optimization, and others. One approach that has seen considerable attention is the use of fairness-related regularization as a component of the loss function used to learn latent factors in a factorization model. For example, Kamishima et al. (2018) proposes adding an independence term to the loss function that penalizes any correlations between the sensitive attribute and the predicted ratings for items; a similar term can also be added to achieve consumer-side fairness (Kamishima and Akaho 2017). Similar approaches can be found in learning-to-rank settings: Beutel et al. (2017) add a penalty term to their pairwise ranking loss function, to ensure that the difference between the ranking scores of the relevant and irrelevant items is uncorrelated with the relevant item's sensitive attribute. It is also possible to directly optimize a learning-to-rank objective to minimize diversity in expected exposure (Diaz et al. 2020).

Although these techniques are relatively common for consumer-side fairness, we believe that caution is required. In domains without subtractibility, there is no interaction between the quality of recommendations delivered to one group versus another. If user group A is well served by the system and group B is poorly served, fairness can be achieved by degrading the experience of group A to match group B, but this means that the system works less well for all. In such a context, putting consumer fairness into the overall loss function is effectively asking the system to make this kind of bargain. Outside of a legal context where fairness across groups may be mandated, this does not seem like the right incentive to build into the learning process. A better practice may be to prioritize the needs of group B in other ways: for example, in preprocessing as noted above, or in other aspects of system design and implementation.

From the point of view of the provider, recommendation is inherently subtractible: the set of entries on a recommendation list is generally fixed and a slot that goes to one item is not available to another. To improve provider-side fairness, Mehrotra et al. (2018) propose trading off consumer relevance and provider fairness via interpolation, probabilistic, and constrained optimization-based recommendation policies within a contextual bandit framework. They find that considering user's affinity to provider fairness improves supplier exposure without severely impacting user satisfaction, which suggests user level heterogeneity in the impact of such interventions across user segments. This work was extended in Mehrotra, Xue, and Lalmas (2020) to incorporate fairness across multiple objectives.

Post-processing

Post-processing approaches focus on modifying the outputs of algorithms to satisfy a fairness criterion. In these methods, fairness constraints do not contribute to the learned objective function for the recommendation model itself, rather they intervene after the output is produced. In recommender systems, this is most commonly described as a *re-ranking* method applied to the original ranked outputs of the recommender.

There are two main approaches to re-ranking: (a) those that treat the problem as a global optimization task and try to improve fairness with respect to an entire set of recommendation lists, and (b) those methods that focus on the fairness of individual lists. An example of the first approach is Sürer, Burke, and Malthouse (2018) that proposes a constrained optimization-based method to enhance fairness (item exposure) for multiple provider groups, avoid unfairness towards under-represented groups and ensure a minimum degree of diversity for consumers. Other optimization approaches, including network flow (Mansoury et al. 2020) and fair allocation (Patro et al. 2020), have been used for similar purposes. These methods impose fairness constraints over the whole set of recommendations at once and are therefore useful where recommendations are generated as part of a batch process, such as push recommendations sent to a group of users all at the same time.

A more common approach is to re-rank individual lists as they are generated. Such approaches are often based on topic diversification models originating in information retrieval (Carbonell and Goldstein 1998; Santos et al. 2010). These methods use a greedy list expansion approach, where the re-ranked list is generated by incrementally adding new items that satisfy a fairness criteria. These approaches also provide the benefit of controlling the balance between the accuracy and the fairness goal. Modani et al. (2017) use a re-ranking approach to enhance provider exposure while preserving relevance. Geyik, Ambler, and Kenthapadi (2019) use a similar greedy approach to produce rankings of job candidates that have a fair distribution of their demographic attributes. A different of optimization is used in Zehlike et al. (2017), which applies the *A** search algorithm to achieve fairness in a ranked list at depth *k*.

In the case of Kiva, Liu et al. (2019) studied a re-ranking technique that supported user-sensitive weighting of the fairness objective in re-ranking so that users with greater affinity or tolerance for a wide range of results encountered more of these results. This work was extended in Sonboli et al. (2020), which consider multiple fairness concerns and users' preferences across multiple concerns. Aligned with the work presented in Tomasi et al. (2020), Sonboli et al. (2020) sees users' propensities towards diversity (measured by entropy) in their recommendation lists as opportunities to increase the exposure of under-represented loans.

Overall, re-ranking approaches offer several advantages. First, the trade-off between accuracy and fairness can be tuned without re-learning the recommendation model (as would be required in other methods). Second, researchers have found that re-ranking can sometimes achieve better trade-offs against accuracy than in-processing models (Abdollahpouri, Burke, and Mobasher 2019).

CONCLUSION AND FUTURE DIRECTIONS

Recommender systems are among the most widely used machine learning applications, and serve a substantial gatekeeping function in social networking and media streaming applications. It is important, therefore, to address their potential for unfair outcomes, both for consumers and providers.

While we have concentrated on distributional harms in this article, that is not to deny the importance of representational harms. Since users treat the output of information access systems as generally representative, even when they are personalized, there is a danger that unbalanced recommendations give a distorted view. This type of harm was highlighted in (non-personalized) search settings in Noble (2018). Also, it should be noted that fairness alone is not a sufficient consideration for creating humane sociotechnical systems. The Belmont Report on human subjects research categorizes justice and beneficence as distinct principles (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). For example, a malevolent system may fairly distribute its harms, but that fairness property does not mitigate the harms themselves, as ominously, if sarcastically, noted in Keyes, Hutson, and Durbin (2019).

We have concentrated here on remedies that are algorithmic or data-centric in nature. However, these are not the only possible avenues for correcting unfair outcomes in a recommendation application. Improvements in system design and enhanced data collection may produce better results even if the core algorithms are unchanged. Another design strategy is to differentiate between the contexts and their associated user expectations, to identify places in the application environment where users may be more amenable to exposure to diverse content (Hansen et al. 2021; Tomasi et al. 2020).

While we have outlined current work in fairness and recommendation, much remains to be studied. There has been relatively little exploration of contexts in which items have inherent utility. As a result, the study of consumer-side fairness has concentrated on differences in performance measures as experienced by different groups of users. There has been more diversity of approaches on the provider side, but fairness for providers seeking to target particular users has not been explored. One reason for these lacunae is the lack of applications and associated data available to academic researchers.

In general, addressing fairness problems with real social impact requires access to data that allows for the assessment of impact on marginalized groups. However, detailed individual demographic information of the type that would be required for this work is precisely the type of data that is most sensitive and most likely to be tightly controlled. We will need to see new forms of collaboration between data holders and recommender systems researchers in order to ensure continued progress in these important areas.

In this article, we have concentrated on examples in which each side of the recommendation interaction has only a single operative fairness concern and all concerns are represented in a single model. However, some practical settings require that multiple, intersecting fairness concerns be considered. Despite some efforts in this area (Sonboli et al. 2020; Wang et al. 2021), the challenge of integrating multiple fairness concerns from multiple stakeholders or from a multi-model recommender system into a coherent and tractable set of fairness objectives is one that needs much further research.

Recommender systems are pervasive and impactful and their social impacts, including the propagation of bias, deserve careful study. This article has highlighted some of the complexities of machine learning fairness as applied to recommendation. In particular, we have shown how the multisided nature of many recommendation platforms gives rise to multifaceted fairness concerns, which are not typically considered in machine learning fairness research.

ACKNOWLEDGMENTS

Authors Sonboli and Burke would like to acknowledge the support of the National Science Foundation under Grant IIS-1911025.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

ORCID

Nasim Sonboli https://orcid.org/0000-0002-6988-7397
Robin Burke https://orcid.org/0000-0001-5766-6434
Michael Ekstrand https://orcid.org/0000-0003-2467-0108
Rishabh Mehrotra https://orcid.org/0000-0002-0836-4605

ENDNOTES

[1] Consumer-side fairness has not arisen so far as a concern in this application.

[2] There also seems to be room for more continuous views of these perspectives, where WAE is treated as a strong prior instead of a logical axiom.

REFERENCES

Abdollahpouri, H. 2020. "Popularity Bias in Recommendation: A Multi-stakeholder Perspective." PhD thesis, University of Colorado.

Abdollahpouri, H., G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato. 2020. "Multistakeholder Recommendation: Survey and Research Directions." *User Modeling and User-Adapted Interaction* 30(1): 127–58.

Abdollahpouri, H., R. Burke, and B. Mobasher. 2019. "Managing Popularity Bias in Recommender Systems with Personalized Re-ranking. " In *Proceedings of the Thirty-Second International Flairs Conference*.

Aggarwal, C. C. 2016. "Model-based Collaborative Filtering." In *Recommender Systems*, 71–138. New York: Springer.

Barbieri, N., and G. Manco. 2011. "An Analysis of Probabilistic Methods for Top-n Recommendation in Collaborative Filtering." In *Machine Learning and Knowledge Discovery in Databases*, edited by D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, 172–87. Berlin, Heidelberg: Springer.

Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104: 671.

Becker, C. D., and E. Ostrom. (1995, November). "Human Ecology And Resource Sustainability: The importance of institutional diversity." *Annual Review of Ecology and Systematics* 26(1): 113–33.

Benjamin, R. (2019, 12). "Race After Technology: Abolitionist Tools for the New Jim Code." *Social Forces* 98(4): 1–3.

Beutel, A., J. Chen, Z. Zhao, and E. H. Chi. 2017. "Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations." *In Proceedings of the Workshop on Fairness, Accountability and Transparency in Machine Learning (FATML), Halifax, Nova Scotia. arXiv preprint arXiv:1707.00075*.

Biega, A. J., K. P. Gummadi, and G. Weikum. 2018. "Equity of Attention: Amortizing Individual Fairness in Rankings." In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR'18*, 405–14 New York, NY, USA: ACM.

Borisyuk, F., L. Zhang, and K. Kenthapadi. 2017. "LiJAR: A System for Job Application Redistribution towards Efficient Career Marketplace." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'17*, 1397–406. New York, NY, USA: ACM. August.

Bozdag, E. (2013, September). "Bias in Algorithmic Filtering and Personalization." *Ethics and Information Technology* 15(3): 209–27.

Burke, R. 2017. "Multisided Fairness for Recommendation." In *Proceedings of the Workshop on Fairness, Accountability and Transparency in Machine Learning (FATML)*, Halifax, Nova Scotia. arXiv:1707.00093.

Carbonell, J., and J. Goldstein. 1998. "The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries." In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'98*, 335–6. New York, NY, USA: ACM.

Chen, I. Y., F. D. Johansson, and D. Sontag. 2018. "Why is My Classifier Discriminatory?" In *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18*, 3543–54. Red Hook, NY, USA: Curran Associates Inc.

Choo, J., C. Lee, D. Lee, H. Zha, and H. Park. 2014. "Understanding and Promoting Micro-finance Activities in kiva.org." In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM'14*, 583–92. New York, NY, USA: ACM.

Diaz, F., B. Mitra, M. D. Ekstrand, A. J. Biega, and B. Carterette. 2020. "Evaluating Stochastic Rankings with Expected Exposure." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM'20*, 275–84 New York, NY, USA: ACM.

Dodge, J., Q. V. Liao, Y. Zhang, R. K. E. Bellamy, and C. Dugan. 2019. "Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment." In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI'19*, 275–85 New York, NY, USA: ACM.

Dwork, C., M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. 2012. "Fairness through Awareness." In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS'12*, 214–26 New York, NY, USA: ACM.

Ekstrand, M., A. Das, R. Burke, and F. Diaz. (To appear). "Fairness and Discrimination in Information Access Systems." *Foundations and Trends in Information Retrieval*, DOI: 10.1561/1500000079. To appear.

Ekstrand, M. D., M. Tian, I. M. Azpiazu, J. D. Ekstrand, O. Anuyah, D. McNeill, and M. S. Pera. 2018. "All the Cool Kids, How Do They Fit In?: Popularity and Demographic Biases in Recommender Evaluation and Effectiveness." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Volume 81 of Proceedings of Machine Learning Research*, ed. S. A. Friedler and C. Wilson, 172–86 New York, NY, USA: PMLR. 23–24 Feb.

Ekstrand, M. D., M. Tian, M. R. I. Kazi, H. Mehrpouyan, and D. Kluver. 2018. "Exploring Author Gender in Book Rating and Recommendation." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys'18*, 242–50 New York, NY, USA: ACM.

Evans, D. S., and R. Schmalensee. 2016. "*Matchmakers: The New Economics of Multisided Platforms*." Boston, MA: Harvard Business Review Press.

Ferraro, A., X. Serra, and C. Bauer. 2021. "Break the Loop: Gender Imbalance in Music Recommenders." In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, CHIIR'21*, 249–54 New York, NY, USA: ACM. March.

Friedler, S. A., C. Scheidegger, and S. Venkatasubramanian. 2021. "The (Im)possibility of Fairness: Different Value Systems Require Different Mechanisms for Fair Decision Making." *Communications of the ACM* 64(4): 136–43. March.

Geyik, S. C., S. Ambler, and K. Kenthapadi. 2019. "Fairness-aware Ranking in Search & Recommendation Systems with Application to Linkedin Talent Search." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD'19*, 2221–31 New York, NY, USA: ACM.

Hansen, C., R. Mehrotra, C. Hansen, B. Brost, L. Maystre, and M. Lalmas. 2021. "Shifting Consumption towards Diverse Content on Music Streaming Platforms." In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM'21*, 238–46 New York, NY, USA: ACM.

Harrison, G., J. Hanson, C. Jacinto, J. Ramirez, and B. Ur. 2020. "An Empirical Study on the Perceived Fairness of Realistic, Imperfect Machine Learning Models." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT*'20*, 392–402 New York, NY, USA: ACM.

Kamishima, T., and S. Akaho. 2017. "Considerations on Recommendation Independence for a Find-good-items Task. " In *Proceedings of the Workshop on Responsible Recommendation (FAccTRec)*.

Kamishima, T., S. Akaho, H. Asoh, and J. Sakuma. 2018. "Recommendation Independence." In *Proceedings of the Conference on Fairness, Accountability and Transparency, Volume 81 of Proceedings of Machine Learning Research*, eds. S. A. Friedler and C. Wilson, 187–201 New York, NY, USA: PMLR.

Karako, C., and P. Manggala. 2018. "Using Image Fairness Representations in Diversity-based Re-ranking for Recommendations." In *Proceedings of the Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP'18*, 23–8 New York, NY, USA: ACM.

Kasinidou, M., S. Kleanthous, P. Barlas, and J. Otterbacher. 2021. "I Agree with the Decision, but They Didn't Deserve This: Future Developers' Perception of Fairness in Algorithmic Decisions." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT'21*, 690–700 New York, NY, USA: ACM.

Kenthapadi, K., B. Le, and G. Venkataraman. 2017. "Personalized Job Recommendation System at Linkedin: Practical Challenges and Lessons Learned." In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys'17*, 346–7 New York, NY, USA: ACM.

Keyes, O., J. Hutson, and M. Durbin. 2019. "A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-nutrient Slurry." In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA'19*, 1–11 New York, NY, USA: ACM.

Kouki, P., S. Fakhraei, J. Foulds, M. Eirinaki, and L. Getoor. 2015. "Hyper: A Flexible and Extensible Probabilistic Framework for Hybrid Recommender Systems." In *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys'15*, 99–106 New York, NY, USA: ACM.

Lee, M. K., D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, D. See, et al. 2019, November. "WeBuildAI: Participatory Framework for Algorithmic Governance." *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–35.

Liu, W., J. Guo, N. Sonboli, R. Burke, and S. Zhang. 2019. "Personalized Fairness-aware Re-ranking for Microlending." In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys'19*, 467–71 New York, NY, USA: ACM.

Mansoury, M., H. Abdollahpouri, M. Pechenizkiy, B. Mobasher, and R. Burke. 2020. "Fairmatch: A Graph-based Approach for Improving Aggregate Diversity in Recommender Systems." In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP'20*, 154–62 New York, NY, USA: ACM.

Mehrotra, R., J. McInerney, H. Bouchard, M. Lalmas, and F. Diaz. 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, CIKM'18*, 2243–51 New York, NY, USA: ACM.

Mehrotra, R., N. Xue, and M. Lalmas. 2020. "Bandit based Optimization of Multiple Objectives on a Music Streaming Platform." In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD'20*, 3224–33 New York, NY, USA: ACM.

Mitchell, S., E. Potash, S. Barocas, A. D'Amour, and K. Lum. 2021. "Algorithmic Fairness: Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application* 8: 141–63.

Modani, N., D. Jain, U. Soni, G. K. Gupta, and P. Agarwal. 2017. "Fairness Aware Recommendations on Behance." In *Advances in Knowledge Discovery and Data Mining*, edited by J. Kim, K. Shim, L. Cao, J.-G. Lee, X. Lin, and Y.-S. Moon, 144–55. Cham: Springer International Publishing.

Mulligan, D. K., J. A. Kroll, N. Kohli, and R. Y. Wong. 2019. "This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology." *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–36.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research." Washington, DC: United States Government.

Ning, X., C. Desrosiers, and G. Karypis. 2015. "A Comprehensive Survey of Neighborhood-based Recommendation Methods." In *Recommender Systems Handbook*, edited by F. Ricci, L. Rokach, and B. Shapira, 37–76. Boston, MA: Springer US.

Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Patro, G. K., A. Biswas, N. Ganguly, K. P. Gummadi, and A. Chakraborty. 2020. "Fairrec: Two-sided Fairness for Personalized Recommendations in Two-sided Platforms." In *Proceedings of The Web Conference 2020, WWW'20*, 1194–204 New York, NY, USA: ACM.

Pedreshi, D., S. Ruggieri, and F. Turini. 2008. "Discrimination-aware Data Mining." In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'08*, 560–8 New York, NY, USA: ACM.

Ricci, F., L. Rokach, and B. Shapira. 2011. "Introduction to Recommender Systems Handbook." In *Recommender Systems Handbook*, edited by F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, 1–35. Boston, MA: Springer US.

Ricci, G., M. de Gemmis, and G. Semeraro. 2012. "Matrix and Tensor Factorization Techniques Applied to Recommender Systems: A Survey." *Matrix* 1(01): 94–8.

Rochet, J.-C., and J. Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association* 1(4): 990–1029.

Roth, A. E., and M. Sotomayor. 1992. "Two-sided matching." *Handbook of Game Theory with Economic Applications* 1: 485–541.

Santos, R. L. T., J. Peng, C. Macdonald, and I. Ounis. 2010. "Explicit Search Result Diversification through Sub-queries." In *Advances in Information Retrieval*, edited by C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. Rüger, and K. van Rijsbergen, 87–99. Berlin, Heidelberg: Springer.

Sapiezynski, P., W. Zeng, R E Robertson, A. Mislove, and C. Wilson. 2019. "Quantifying the Impact of User Attention on Fair Group Representation in Ranked Lists." In *Companion Proceedings of the 2019 World Wide Web Conference, WWW'19*, 553–62 New York, NY, USA: ACM.

Selbst, A. D., D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi. 2019. "Fairness and Abstraction in Sociotechnical

Systems." In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT*19*, 59–68 New York, NY, USA: ACM.

Singh, A., and T. Joachims. 2018. "Fairness of Exposure in Rankings." In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD'18*, 2219–28 New York, NY, USA: ACM.

Sonboli, N., F. Eskandanian, R. Burke, W. Liu, and B. Mobasher. 2020. "Opportunistic Multi-aspect Fairness through Personalized Re-ranking." In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP'20*, 239–47 New York, NY, USA: ACM.

Sonboli, N., J. J. Smith, F. Cabral Berenfus, R. Burke, and C. Fiesler. 2021. "Fairness and Transparency in Recommendation: The Users' Perspective." In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP'21*, 274–9 New York, NY, USA: ACM.

Srivastava, M., H. Heidari, and A. Krause. 2019. "Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD'19*, 2459–68 New York, NY, USA: ACM.

Steck, H. 2018. "Calibrated Recommendations." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys'18*, 154–62 New York, NY, USA: ACM.

Sürer, O., R. Burke, and E. C. Malthouse. 2018. "Multistakeholder Recommendation with Provider Constraints." In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys'18*, 54–62 New York, NY, USA: ACM.

Tomasi, F., R. Mehrotra, A. Pappu, J. Bütepage, B. Brost, H. Galvão, and M. Lalmas. 2020. "Query Understanding for Surfacing Underserved Music Content." In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM'20*, 2765–72 New York, NY, USA: ACM.

Wang, J., and S. Yuan. 2015. "Real-time Bidding: A New Frontier of Computational Advertising Research." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM'15*, 415–6 New York, NY, USA: ACM.

Wang, X., N. Thain, A. Sinha, F. Prost, E. H. Chi, J. Chen, and A. Beutel. 2021. "Practical Compositional Fairness: Understanding Fairness in Multi-component Recommender Systems." In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM'21*, 436–44 New York, NY, USA: ACM.

Xia, B., J. Yin, J. Xu, and Y. Li. 2019. "WE-Rec: A Fairness-aware Reciprocal Recommendation based on Walrasian Equilibrium." *Knowledge-Based Systems* 182:104857.

Yang, K., and J. Stoyanovich. 2017. "Measuring Fairness in Ranked Outputs." In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management, SSDBM'17*, New York, NY, USA: ACM.

Yao, S., and B. Huang. 2017. "Beyond Parity: Fairness Objectives for Collaborative Filtering." In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, 2925–34 Red Hook, NY, USA: Curran Associates Inc.

Zehlike, M., F. Bonchi, C. Castillo, S. Hajian, M. Megahed, and R. Baeza-Yates. 2017. "FA*IR: A Fair Top-k Ranking Algorithm." In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1569–78. New York, NY, USA: ACM.

Zhang, S., L. Yao, A. Sun, and Y. Tay. (2019, February). "Deep Learning based Recommender System: A Survey and New Perspectives." *ACM Computing Surveys (CSUR)* 52(1): 1–38.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

AUTHOR BIOGRAPHIES

Nasim Sonboli is a Computing Innovation Fellow at Tufts University, where she researches the intersection of Responsible Recommender Systems and General Data Protection Regulations (GDPR). She got her Ph.D. from the University of Colorado, Boulder, in 2022, where she worked on fairness-aware recommender systems under the supervision of Dr. Robin Burke.

Robin Burke is a Professor of Information Science at the University of Colorado, Boulder. His research group, That Recommender Systems Lab, conducts research in personalized recommender systems, including fairness, accountability, and transparency in recommendation through the integration of objectives from diverse stakeholders.

Michael D. Ekstrand is an Associate Professor of Computer Science at Boise State University, where he co-directs the People and Information Research Team. His 15 years of research on recommender systems, information retrieval, and human–computer interaction aim to ensure that information access technologies are built and deployed in ways that promote human flourishing.

Rishabh Mehrotra currently works as a Director of Machine Learning at ShareChat, London. Prior to ShareChat, he was a Staff Scientist and Engineer at Spotify where he led multiple recommendation products from basic research to production across 400+ million users. His current research focuses on machine learning for marketplaces, multi-objective modeling of recommenders and creator ecosystem.

How to cite this article: Sonboli, N., R. Burke, M. Ekstrand, and R. Mehrotra. 2022. "The multisided complexity of fairness in recommender systems." *AI Magazine* 43: 164–76. https://doi.org/10.1002/aaai.12054