



# Fairness in Music Recommender Systems: A Stakeholder-Centered Mini Review

Karlijn Dinnissen\* and Christine Bauer

Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

The performance of recommender systems highly impacts both music streaming platform users and the artists providing music. As fairness is a fundamental value of human life, there is increasing pressure for these algorithmic decision-making processes to be fair as well. However, many factors make recommender systems prone to biases, resulting in unfair outcomes. Furthermore, several stakeholders are involved, who may all have distinct needs requiring different fairness considerations. While there is an increasing interest in research on recommender system fairness in general, the music domain has received relatively little attention. This mini review, therefore, outlines current literature on music recommender system fairness from the perspective of each relevant stakeholder and the stakeholders combined. For instance, various works address gender fairness: one line of research compares differences in recommendation quality across user gender groups, and another line focuses on the imbalanced representation of artist gender in the recommendations. In addition to gender, popularity bias is frequently addressed; yet, primarily from the user perspective and rarely addressing how it impacts the representation of artists. Overall, this narrative literature review shows that the large majority of works analyze the current situation of fairness in music recommender systems, whereas only a few works propose approaches to improve it. This is, thus, a promising direction for future research.

## OPEN ACCESS

### Edited by:

Lina Yao,  
University of New South Wales,  
Australia

### Reviewed by:

Mirko Marras,  
University of Cagliari, Italy

### \*Correspondence:

Karlijn Dinnissen  
k.dinnissen@uu.nl

### Specialty section:

This article was submitted to  
Recommender Systems,  
a section of the journal  
Frontiers in Big Data

Received: 05 April 2022

Accepted: 15 June 2022

Published: 22 July 2022

### Citation:

Dinnissen K and Bauer C (2022)  
Fairness in Music Recommender  
Systems: A Stakeholder-Centered  
Mini Review.  
Front. Big Data 5:913608.  
doi: 10.3389/fdata.2022.913608

**Keywords:** bias mitigation, fairness, music recommendation systems, stakeholders, literature review

## 1. INTRODUCTION

The art of music recommendation was traditionally performed exclusively by people, such as DJs, record store owners, and friends. In the last few decades, however, this task has been partially automated using machine learning (ML) techniques; recommender systems (RSs) in particular (Celma, 2010b). Learning from large-scale user behavior and music features, so-called music recommender systems (MRSs) can automatically produce recommendations tailored to a specific user (Ekstrand et al., 2022). This is one of the reasons why music streaming platforms, that typically integrate MRSs, have become one of the main sources of music consumption (IFPI, 2020). Consequently, the performance of MRSs highly impacts users' overall music listening experience (Lee et al., 2019) and considerably impacts artists in terms of exposure and resulting royalty payments (Ferraro et al., 2021b).

ML system users frequently perceive RS decisions as objective (Helberger et al., 2020). However, many factors make such systems' processes prone to biases, resulting in unfair outcomes (Ekstrand et al., 2022). One such factor is that ML models are created and trained by humans whose intrinsic

biases may be carried over. Furthermore, the data that is used to train ML models may contain biases as well. This is problematic, as fairness is a fundamental value of human life (Folger and Cropanzano, 1998; Tyler and Smith, 1998). Moreover, anti-discrimination regulations explicitly prohibit that characteristics such as gender, age, and nationality cause different outcomes for otherwise similar people (Civil Rights Act, 1964; Age Discrimination in Employment Act, 1967; European Union, 2010, Art. 21). It is, therefore, crucial to critically review MRSs for any form of unfairness to ensure that they do not unfairly disadvantage any user or artist.

Overall, there is an increasing interest in research on fairness in ML in general (Hutchinson and Mitchell, 2019), and in RSs in particular (Ekstrand et al., 2019). One of the challenges in fairness research is that it is scattered across several disciplines (Holstein et al., 2019; Selbst et al., 2019). Moreover, it concerns several stakeholders with distinct fairness needs, calling for various bias mitigation strategies (Ekstrand et al., 2022). Considering those needs is, thus, key to both, understanding fairness in music recommendation algorithms and designing strategies to improve it. To the best of our knowledge, an overview of such needs and strategies does not yet exist for the music recommendation field specifically. Therefore, this work addresses the following research question: *What is the state-of-the-art of MRS fairness research from the various stakeholders' perspectives?* To address this RQ, we conduct a narrative literature review, giving a thorough overview of works that *explicitly target RS fairness in the music domain*. We also include some works that are not explicitly concerned with fairness, yet address fairness as a side effect.

In Section 2, we first define each relevant stakeholder group. Then, in the Sections 2.1, 2.2, and 2.3, we present our narrative literature review in which we address each of the relevant stakeholders separately. In Section 3, we conclude this work with a discussion of the lessons learned from this overview and derive research gaps, thereby forming a solid basis for future research.

## 2. FAIRNESS FOR MULTIPLE STAKEHOLDERS IN MUSIC RECOMMENDER SYSTEMS

The digital music value chain embraces a wide set of stakeholders, who have different goals and interests regarding the music being recommended (Bauer and Zangerle, 2019). Recommender systems literature typically distinguishes three stakeholders: platform users (end consumers), item providers, and the platform itself (Abdollahpouri et al., 2017b; Burke, 2017; Sonboli et al., 2021). Some variations can be found in literature; for instance, Mehrotra et al. (2018) and Patro et al. (2020) only consider user and item provider as stakeholders, yet not the platform; conversely, Jannach and Bauer (2020) include society at large as a fourth stakeholder.

In MRSs, there are three main stakeholders. Firstly, the **users** (Section 2.1)—also called consumers or customers—are the party consuming the music recommendations. A user may be an individual or a group of individuals, served by music streaming platforms. As individuals have different profiles containing, for

instance, different characteristics, preferences, or needs, MRSs might create a better experience for some user groups than for others. Ideally, a MRS creates a good user experience for all users.

Secondly, the **item providers** (Section 2.2)—also referred to as producers or suppliers—form the stakeholder supplying the recommended music and benefiting from it being consumed or purchased. In MRS research, the artists (including performers, music producers, and songwriters) are typically the item providers, but record companies or publishers representing several artists may also be considered item providers. Each item provider usually represents a multitude of items in the form of music tracks. A higher MRS ranking for an item implies a higher chance of exposure to users, resulting in a higher chance that users interact with the item (Biega et al., 2018; Diaz et al., 2020). This is desirable, as item interaction results in revenue (Deldjoo et al., 2021). Typically, item providers have little control over when and to whom their items are recommended (Burke, 2017; Ferraro et al., 2021b).

Thirdly, the **platform** exists at the center of the music recommender ecosystem (Abdollahpouri and Essinger, 2017; Smets et al., 2022). Music streaming platforms (such as Apple Music, Deezer, Pandora, QQ Music, Spotify, and Tidal) act as an interface between huge repositories of music tracks and millions of music consumers. On such platforms, the interaction between users and items is facilitated by a MRS. A platform needs to attract and retain both users as well as item providers and, thus, benefits from a successful match between users and items (Burke, 2017). As the platforms are in control of the MRS they embed (Bauer and Zangerle, 2018) and can even significantly influence consumption decisions through functionalities such as curated playlists (Aguilar and Waldfogel, 2021), they are typically not considered being at risk of unfair treatment. Rather, platforms might impose fairness constraints to satisfy an organizational mission or meet demands of, e.g., government regulators or interest groups (Ekstrand et al., 2022). Further, there is increasing external pressure to make these platforms and their integrated MRSs fairer (Burke et al., 2018; Bauer and Zangerle, 2019; Patro et al., 2020; Ferraro et al., 2021b; Melchiorre et al., 2021).

As multiple stakeholders with possibly diverging interests are involved and affected by MRSs, **multi-stakeholder** research (Section 2.3) addresses several stakeholder groups simultaneously. Each stakeholder may have distinct fairness needs, which may further differ per context and application (Burke, 2017; Ekstrand and Kluver, 2021). Consequently, solely optimizing RSs on metrics such as user satisfaction may be detrimental to user fairness, item provider fairness, or both (Bauer and Zangerle, 2019; Patro et al., 2020). Hence, several studies urge to consider the interests of all stakeholder groups (Burke, 2017; Mehrotra et al., 2018, 2020). We note that research that addresses fairness, for example, for item providers, while also measuring performance indicators such as user satisfaction in the evaluation, are not necessarily multi-stakeholder approaches; a multi-stakeholder perspective integrates the various stakeholders fundamentally.

**Table 1** provides an overview of the papers on fairness in MRSs considered in this narrative literature review. It

**TABLE 1** | Overview of literature on fairness in music recommender systems.

References	Improvement focus	Methodology	Topic	Considered fairness attribute(s)	Stakeholder focus	Dataset source
Bauer et al. (2017)		Conceptual, interview	Negative impact for non-superstar artists	Popularity	Item provider	–
Bauer and Schedl (2018)	x	Data analysis, offline experiment	Improving accuracy by considering mainstreamness and country	User country, user “mainstreamness”	User	LFM-1b
Bauer and Schedl (2019)	x	Data analysis, offline experiment	Improving accuracy by considering mainstreamness and country	User country, user “mainstreamness”	User	LFM-1b
Boratto et al. (2022)	x (reproduction)	Systematic literature review, reproduction	Reproducing and comparing unfairness mitigation strategies	User age, user gender	User	LFM-1K
Celma (2010b)		Data analysis	Promotion of niche items	Popularity	User	Proprietary (Last.fm and MySpace)
Celma and Cano (2008)		Data analysis, offline experiment	Investigating popularity bias in collaborative filtering	Popularity	User	Proprietary (Last.fm)
Ekstrand et al. (2018)		Data analysis, offline experiment	Recommender effectiveness across demographics and popularity levels	Popularity, user age, user gender	User	LFM-1K, LFM-360K
Epps-Darling et al. (2020)		Data analysis	Analysis of gender distribution across popularity levels	Artist gender, popularity	Item provider	Proprietary (Spotify)
Ferraro et al. (2020)		Offline experiment	Evaluating influence of recommendation bias on artist exposure	Contemporaneity, country, gender, type (all artist attributes)	Item provider	LFM-360K
Ferraro et al. (2021a)	x	Interviews, data analysis, offline experiment, long-term simulation	Improving gender fairness	Artist gender	Item provider	LFM-360K, LFM-1b
Ferraro et al. (2021b)		Interviews	Impact of recommender systems on artists	Age, contemporaneity, country, diversity, gender, popularity (all artist attributes)	Item provider	–
Flexer et al. (2018)		Data analysis	Hubness as a technical algorithmic bias in high dimensional machine learning	<sup>–a</sup>	User, item provider	Proprietary (FM4 SoundPark)
Htun et al. (2021)		User study	Perception of fairness per user personality type	<sup>–b</sup>	User	–
Kowald et al. (2021)		Data analysis, offline experiment	Characteristics of niche music and music listeners	User “mainstreamness”	User	LFM-1b
Kowald et al. (2020)		Data analysis, offline experiment	Investigating the impact of popularity bias on niche items, and users favoring those items	Popularity, user “mainstreamness”	User	LFM-1b

(Continued)

TABLE 1 | Continued

References	Improvement focus	Methodology	Topic	Considered fairness attribute(s)	Stakeholder focus	Dataset source
Lesota et al. (2021)		Data analysis, offline experiment	Effect of popularity bias per gender	Popularity, user gender	User	LFM-2b
Mehrotra et al. (2018)	x	Offline experiment	Relevance, fairness and satisfaction trade-off in a two-sided marketplace	Popularity	User, item provider	Proprietary (Spotify)
Mehrotra et al. (2020)	x	Offline experiment	Contextual bandits that consider multiple objectives (e.g., gender diversity, niche items)	Artist gender, popularity	User, item provider	Proprietary (Spotify), Simulated data
Melchiorre et al. (2021)	x	Data analysis, offline experiment	Improvement of gender fairness considering popularity bias	User gender	User	LFM-2b
Mousavifar and Vassileva (2022)		User study	Using explanations to increase user satisfaction with fair recommendation	Popularity	User, item provider	–
Neophytou et al. (2022)		Offline experiment, reproduction	Reproducing recommendation utility for different user groups	Popularity, user age, user country, user gender	User	LFM-360K
Oliveira et al. (2017)	x	Offline experiment	Considering diversification and user preferences simultaneously in a multi-objective approach	Contemporaneity, gender, genre, locality (all artist attributes)	User, item provider	LFM-1b, Simulated data
Schedl and Bauer (2017)		Offline experiment	Improving accuracy by considering mainstreamness	User "mainstreamness"	User	LFM-1b
Shakespeare et al. (2020)		Data analysis, offline experiment	Investigating gender fairness	Artist gender	Item provider	LFM-360K, LFM-1b, Simulated data

<sup>a</sup>Hubness can create unfairness for any attribute.

<sup>b</sup>Not transparent which fairness attributes participants were considering.

also includes information on the research focus, methodology, considered fairness attributes, the stakeholders in the loop, and the datasets used for conducting the research.

## 2.1. User Perspective

From the user perspective, fairness in MRS is primarily studied based on distinct user groups defined by personal characteristics. In addition to groups based on protected characteristics, groups differentiated by other characteristics may experience unfairness as well.

A wealth of literature analyzes popularity bias and subsequent mitigation strategies in various application domains (e.g., Figueiredo et al., 2014; Abdollahpouri et al., 2017a; Wei et al., 2021). It is, for instance, widely acknowledged that collaborative filtering-based recommendation approaches are prone to popularity bias (Celma and Cano, 2008; Jannach et al.,

2015). The music domain is a well-known example of the long-tail economy (Anderson, 2006) and popularity bias is, thus, particularly relevant. It can be considered either a problem (Anderson, 2006) or a desired feature as popularity in the community signifies some relevancy (Celma, 2010b). In general, many works address popularity bias in MRSs with various intentions. Some address the cold-start problem for items without prior user ratings to make them recommendable (e.g., Ferraro, 2019); others aim at increasing user satisfaction by adding novelty through recommending items from the long tail (e.g., Bedi et al., 2014); yet other works leverage the long tail to specifically address discovery (e.g., Domingues et al., 2013). While fairness is not always necessarily put in the loop of the investigation, this research thread does address fairness aspects.

As for insights from works that explicitly consider user fairness in MRSs, recommendation accuracy tends to be higher

for “mainstream” users, who are inclined toward what is popular, compared to “beyond-mainstream” users who prefer less popular items (Kowald et al., 2020, 2021). This also holds when defining user groups based on a more fine-grained music taste level (Schedl and Bauer, 2017; Kowald et al., 2021). Some works (e.g., Bauer and Schedl, 2019) have proposed mechanisms that better reflect the preferences of beyond-mainstream users.

When defining user groups based on user country, popularity bias also negatively affects MRS performance for groups from countries with preferences beyond the global mainstream (Bauer and Schedl, 2018; Neophytou et al., 2022). In a later work, Bauer and Schedl (2019) propose context-prefiltering approaches to mitigate this issue. Zooming in on another user characteristic, several studies investigate gender. They show that popularity bias particularly affects minority gender groups (in these studies: women), resulting in lower-quality recommendations in terms of accuracy and coverage (e.g., Lesota et al., 2021; Melchiorre et al., 2021). In addition to finding similar results for user gender, Ekstrand et al. (2018) and its reproducibility study by Neophytou et al. (2022) found performance differences for different user age groups, too. Here, the older user group received lower-quality recommendations.

Lastly, on the mitigation side, Boratto et al. (2022) present a reproducibility study focusing on user age and gender, applying various mitigation strategies in the music and movie domains. Different from the movie domain, the size of the user group was not indicative of the recommender accuracy in the music domain. Given their indecisive results, it is important to look beyond popularity bias and demographic group size to understand the drivers of demographic differences.

Melchiorre et al. (2020) define user groups based on personality traits. In contrast to the work on gender, age, and country, personality traits are not among the characteristics acknowledged by anti-discrimination regulations, and fairness research is also not clear about this issue either. Nonetheless, they may be a source of bias and an opportunity for MRS improvement. Melchiorre et al. (2020) illustrate this by showing that scoring low on the personality traits openness, extraversion, and conscientiousness results in higher recommender performance, whereas scoring low on neuroticism or agreeableness leads to lower performance. Additionally, Htun et al. (2021) study the effect of personality traits on the perception of fairness in group recommendations when creating group music playlists. Here, the personality trait openness is negatively correlated with the perception that fairness is important in groups. Given that diversity needs and personality traits correlate (Chen et al., 2013), considering those traits in user modeling may help improve MRS performance.

## 2.2. Item Provider Perspective

When considering harm against music providers caused by unfairness in MRSs, research mainly focuses on group fairness (Singh and Joachims, 2018). Item provider groups in MRS research have been primarily defined based on gender (Ekstrand and Kluver, 2021; Ferraro et al., 2021a). Several approaches are used to study and mitigate item provider gender bias, illustrating that a multifaceted approach is needed.

To date, most research has focused on understanding existing gender biases (e.g., Wang and Horvát, 2019; Epps-Darling et al., 2020). The former analyzed a Spotify streaming sample and found a disparity between artist genders in users’ listening behavior. In “organic” streaming, such as streams originating from a user library or user’s search, 21.75% of tracks were from either a woman or multi-gender formation. For streams programmed by MRSs, this number was 23.55%. This gender gap in listening behavior is further reflected in commonly used datasets such as LFM-1b and LFM-360k, in which 23% of (solo) artists are women (Ferraro et al., 2021a). These datasets roughly reflect the gender gap in business reality (Youngs, 2019; Epps-Darling et al., 2020). Overall, these percentages reflect the barriers to entry, and subsequently climbing to the top, for minority genders. In addition, pre-existing gender biases might influence which tracks users select in a MRS. Ferraro et al. (2020) and Shakespeare et al. (2020) found that collaborative filtering algorithms could propagate or even amplify those biases in a MRS, thereby negatively impacting minority genders. In the latter, no evidence was found for the algorithms introducing new gender biases, which is supported by Epps-Darling et al. (2020) who found that recommendation-based streaming even contained a slightly higher proportion of tracks by women than in organic listening. On the gender bias mitigation side, re-ranking is a promising method. Ferraro et al. (2021a) demonstrate breaking bias amplification through gradually increasing exposure for minority genders.

In addition to gender, Oliveira et al. (2017) consider genre, locality, and contemporaneity. Embracing these attributes, they introduce a multi-objective approach to diversification that addresses fairness for users and item providers alike. Ferraro et al. (2020) use similar categories and add artist type (e.g., solo artist, band). Their analysis of the locality attribute indicates that group size may foster exposure: the artists from the most represented countries in the dataset (here: United Kingdom and United States) reached high exposure, while minority countries were penalized.

Defining item provider groups based on their popularity level has been investigated, too (Celma and Cano, 2008; Bauer et al., 2017). Although popularity bias is a frequently researched topic, fairness goals are predominantly defined for MRS users and not item providers. One exception to this is Flexer et al. (2018) who study the “hubness” phenomenon, which can occur in content-based RS models that use song similarity as their main feature. Hubness refers to some music tracks being connected to many other tracks in the database without a clear semantic musical connection. This may introduce unfairness for tracks that are more similar semantically, but not recommended as often.

To date, one study directly discusses fairness in MRSs with the item providers themselves: Ferraro et al. (2021b) interviewed artists about their perception of fairness in MRSs, and how item provider fairness could be improved on music streaming platforms. In those interviews, the main noted fairness improvement areas relate to nurturing diversity in general, and in particular to gender representation, addressing popularity bias, and providing a better representation of genres beyond the mainstream. These

topics also correspond to the aforementioned research focuses in literature.

### 2.3. Multi-Stakeholder Perspective

Studies may simultaneously take several different MRS stakeholder objectives (e.g., satisfaction, utility, fairness, or diversity) into account. Generally, across application domains, a trade-off between such objectives is reported (Cramer et al., 2018; Mehrotra et al., 2018; Singh and Joachims, 2018), though it is possible that multi-stakeholder objective optimization benefits all stakeholders. Item provider fairness, for example, does not have to be detrimental to user satisfaction (Mehrotra et al., 2018), and persuasive strategies may even be implemented to promote new and less popular artists while increasing user satisfaction (Mousavifar and Vassileva, 2022). Furthermore, even if users do not directly benefit from or even consider fairness for item providers, they indicate that it is important to incorporate it in RSs (Sonboli et al., 2021).

Overall, fairness-related multi-stakeholder MRS work mainly defines objectives and stakeholders rather than aiming to improve fairness. Mehrotra et al. (2018), though, do contribute to fairness improvement by introducing a counterfactual estimation framework that balances provider fairness with user relevance and can optimize either, aiming to provide an alternative for expensive online A/B tests. In another study, Mehrotra et al. (2020) use “contextual bandits” that can optimize multiple objectives simultaneously in a fair way, this time focusing on user- and platform objectives as opposed to item providers.

We might also draw inspiration from multi-stakeholder MRS research where fairness is not an explicitly defined goal. For instance, Unger et al. (2021) introduce a multi-objective RS that aims to fulfill both user satisfaction (measured by saves, likes, and engagement) and item provider satisfaction (determined by, e.g., acquiring new fans). A similar approach may be taken to implement fairness objectives for multiple stakeholders. Patro et al. (2020) propose FairRec, which exhibits fairness for both user and item provider while the loss in overall recommendation quality remains marginal. FairRec has, however, not been applied to the music domain yet.

## 3. DISCUSSION AND CONCLUSIONS

This literature overview demonstrates that, while there is increasing interest in research on fairness in RSs in general, comparatively little research has addressed the music domain. Below, we discuss the main findings we derive from this review.

### 3.1. Research Focus

Contrary to what literature frequently claims (e.g., Patro et al., 2020; Ferraro et al., 2021b), fairness in this context has been addressed from both the user perspective and the item provider perspective. Yet, multi-stakeholder approaches to fairness are scarce. This review also shows that the large majority of MRS fairness works analyzes the current situation, using existing approaches and available datasets. We, therefore,

identify improvement-focused research as the main research gap. A major challenge remains here: we still need to improve our understanding of the normative nature of fairness. While an entirely fair system is likely unachievable, it is crucial to recognize RS fairness issues, mitigate them, and incrementally improve fairness over the current state.

### 3.2. Gender Bias

Interestingly, various MRS works address gender fairness, both for user and item providers. We speculate that this focus has emerged from gender being an immutable characteristic, the wide acknowledgment that gender fairness is of societal relevance, and gender labels being available to some extent in relevant datasets. While it is a known limitation that a binary concept of gender oversimplifies gender expression, current datasets predominantly restrict the gender labels to man and woman (Shakespeare et al., 2020; Ferraro et al., 2021a; Boratto et al., 2022). A notable exception is the work by Epps-Darling et al. (2020).

### 3.3. Popularity Bias

While popularity bias may be considered an item provider fairness issue as the gap between popular and unpopular items increases, research frequently focuses on the user. Addressing popularity is seen as a means to provide more diverse content to increase user satisfaction. Similarly, we observe that some works do not explicitly focus on fairness, but still demonstrate fairness intentions or improvements in their research. As this review focused on works that address fairness *explicitly*, this overview is not intended to be exhaustive.

### 3.4. Data Availability

As can be seen in **Table 1**, the most frequently used datasets originate from Last.fm: LFM-1b (Schedl, 2016), LFM-1K, LFM-360K (both Celma, 2010a), and the recently added LFM-2b (Schedl et al., 2022). This results in only a few datasets being used for research on fairness in MRS; most of which are either based on the same or similar Last.fm data, or are proprietary and therefore not accessible to other researchers. Overall, this means that the used datasets might not be representative. Additionally, only a few open datasets in the music domain contain user interaction or preference data. They also typically include only limited fairness-related stakeholder metadata (e.g., gender, age, ethnicity), as sensitive data is often not shared (Stoikov and Wen, 2021). For ethical reasons, it is debatable whether it should be. Lastly, a current limitation is the focus on short-term bias mitigation, while real world-systems are active over years (Shakespeare et al., 2020). Longitudinal data or simulation frameworks are needed to better address these temporary aspects and to study fairness in MRS in the long run. Summing up, to achieve significant MRSs fairness improvements, richer and more representative data is needed.

## AUTHOR CONTRIBUTIONS

KD and CB contributed to writing and revising the manuscript draft, as well as the final submitted version.

## REFERENCES

- Abdollahpouri, H., Burke, R., and Mobasher, B. (2017a). "Controlling popularity bias in learning-to-rank recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17* (New York, NY: Association for Computing Machinery), 42–46. doi: 10.1145/3109859.3109912
- Abdollahpouri, H., Burke, R., and Mobasher, B. (2017b). "Recommender systems as multistakeholder environments," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17* (New York, NY: Association for Computing Machinery), 347–348. doi: 10.1145/3079628.3079657
- Abdollahpouri, H., and Essinger, S. (2017). "Multiple stakeholders in music recommender systems," in *1st International Workshop on Value-Aware and Multistakeholder Recommendation at RecSys 2017, VAMS '17* (New York, NY: Association for Computing Machinery), 1–3.
- Age Discrimination in Employment Act (1967). *Age Discrimination in Employment Act of 1967*.
- Aguiar, L., and Waldfogel, J. (2021). Platforms, power, and promotion: Evidence from spotify playlists\*. *J. Indus. Econ.* 69, 653–691. doi: 10.1111/joie.12263
- Anderson, C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More*. New York, NY: Hyperion.
- Bauer, C., Kholodylo, M., and Strauss, C. (2017). "Music recommender systems: challenges and opportunities for non-superstar artists," in *30th Bled eConference*, eds A. Pucihar, M. K. Borštnar, C. Kittl, P. Ravesteijn, R. Clarke, and R. Bons (Maribor: University of Maribor Press), 21–32. doi: 10.18690/978-961-286-043-1.3
- Bauer, C., and Schedl, M. (2018). "On the importance of considering country-specific aspects on the online-market: an example of music recommendation considering country-specific mainstream," in *51st Hawaii International Conference on System Sciences, HICSS '18* (Menlo Park, HI), 3647–3656. doi: 10.24251/HICSS.2018.461
- Bauer, C., and Schedl, M. (2019). Global and country-specific mainstreamness measures: definitions, analysis, and usage for improving personalized music recommendation systems. *PLoS ONE* 14:e217389. doi: 10.1371/journal.pone.0217389
- Bauer, C., and Zangerle, E. (2018). "Information imbalance and responsibility in recommender systems," in *2nd Workshop on Green (Responsible, Ethical and Social) IT and IS—the Corporate Perspective (GRES-IT/IS), GRES-IT/IS 2018*, eds B. Krumay and R. Brandtweiner (Vienna: Department for Informationsverarbeitung und Prozessmanagement; WU Vienna University of Economics and Business), 1–3.
- Bauer, C., and Zangerle, E. (2019). "Leveraging multi-method evaluation for multi-stakeholder settings," in *Proceedings of the 1st Workshop on the Impact of Recommender Systems, Co-Located With 13th ACM Conference on Recommender Systems (ACM RecSys 2019), Vol. 2462 of ImpactRS '19*, eds O. S. Shalom, D. Jannach, and I. Guy (New York, NY: Association for Computing Machinery), 1–3. Available online at: <http://ceur-ws.org/Vol-2462/short3.pdf> (accessed June 15, 2022).
- Bedi, P., Gautam, A., Richa, and Sharma, C. (2014). "Using novelty score of unseen items to handle popularity bias in recommender systems," in *2014 International Conference on Contemporary Computing and Informatics, IC3I '14* (Red Hook, NY: Curran Associates, Inc.), 934–939. doi: 10.1109/IC3I.2014.7019608
- Biega, A. J., Gummadi, K. P., and Weikum, G. (2018). "Equity of attention: amortizing individual fairness in rankings," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18* (New York, NY: Association for Computing Machinery), 405–414. doi: 10.1145/3209978.3210063
- Boratto, L., Fenu, G., Marras, M., and Medda, G. (2022). "Consumer fairness in recommender systems: contextualizing definitions and mitigations," in *Advances in Information Retrieval*, eds M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørsvåg, and V. Setty (Cham: Springer International Publishing), 552–566. doi: 10.1007/978-3-030-99736-6\_37
- Burke, R. (2017). "Multisided fairness for recommendation," in *Proceedings of the Workshop on Fairness, Accountability and Transparency in Machine Learning, Held at KDD 2017, FAT/ML '17* (Halifax, NS), 1–5.
- Burke, R., Sonboli, N., and Ordonez-Gauger, A. (2018). "Balanced neighborhoods for multi-sided fairness in recommendation," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Vol. 81 of Proceedings of Machine Learning Research FAT\* '18*, eds S. A. Friedler and C. Wilson (New York, NY: Association for Computing Machinery), 202–214. Available online at: <http://proceedings.mlr.press/v81/burke18a/burke18a.pdf> (accessed June 15, 2022).
- Celma, Ó. (2010a). *Chapter 3: Music Recommendation*. Berlin; Heidelberg: Springer Berlin Heidelberg, 43–85.
- Celma, Ó. (2010b). *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Berlin; Heidelberg: Springer.
- Celma, Ó., and Cano, P. (2008). "From hits to niches? or how popular artists can bias music recommendation and discovery," in *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition, NETFLIX '08* (New York, NY: Association for Computing Machinery), 1–8. doi: 10.1145/1722149.1722154
- Chen, L., Wu, W., and He, L. (2013). "How personality influences users' needs for recommendation diversity?," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13* (New York, NY: Association for Computing Machinery), 829–834. doi: 10.1145/2468356.2468505
- Civil Rights Act (1964). *Civil Rights Act of 1964. Title VII, Equal Employment Opportunities*. Civil Rights Act.
- Cramer, H., Garcia-Gathright, J., Springer, A., and Reddy, S. (2018). Assessing and addressing algorithmic bias in practice. *Interactions* 25, 58–63. doi: 10.1145/3278156
- Deldjoo, Y., Anelli, V. W., Zamani, H., Bellogón, A., and Di Noia, T. (2021). A flexible framework for evaluating user and item fairness in recommender systems. *User Model. User Adapt. Interact.* 31, 457–511. doi: 10.1007/s11257-020-09285-1
- Diaz, F., Mitra, B., Ekstrand, M. D., Biega, A. J., and Carterette, B. (2020). "Evaluating stochastic rankings with expected exposure," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20* (New York, NY: Association for Computing Machinery), 275–284. doi: 10.1145/3340531.3411962
- Domingues, M. A., Gouyon, F., Jorge, A. M., Leal, J., Vinagre, J., Lemos, L., and Sordo, M. (2013). Combining usage and content in an online recommendation system for music in the long tail. *Int. J. Multim. Inform. Retrieval* 2, 3–13. doi: 10.1007/s13735-012-0025-1
- Ekstrand, M. D., Burke, R., and Diaz, F. (2019). "Fairness and discrimination in recommendation and retrieval," in *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY: Association for Computing Machinery), 576–577. doi: 10.1145/3298689.3346964
- Ekstrand, M. D., Das, A., Burke, R., and Diaz, F. (2022). Fairness in information access systems (to appear). *Found. Trends Inform. Retrieval*. 1–92. doi: 10.48550/arXiv.2105.05779. Available online at: <https://arxiv.org/abs/2105.05779> (accessed June 15, 2022).
- Ekstrand, M. D., and Klüber, D. (2021). Exploring author gender in book rating and recommendation. *User Model. User Adapt. Interact.* 31, 377–420. doi: 10.1007/s11257-020-09284-2
- Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., et al. (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*, eds S. A. Friedler and C. Wilson (New York, NY: Association for Computing Machinery), 172–186. Available online at: <http://proceedings.mlr.press/v81/ekstrand18b/ekstrand18b.pdf> (accessed June 15, 2022).
- Epps-Darling, A., Takeo Bouyer, R., and Cramer, H. (2020). "Artist gender representation in music streaming," in *Proceedings of the 21st International Society for Music Information Retrieval Conference, ISMIR '20* (Montréal, QC), 248–254. Available online at: <https://archives.ismir.net/ismir2020/paper/000148.pdf> (accessed June 15, 2022).
- European Union (2010). *Charter of Fundamental Rights of the European Union, Vol. 53*. Brussels: European Union.
- Ferraro, A. (2019). "Music cold-start and long-tail recommendation: bias in deep representations," in *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys '19* (New York, NY: Association for Computing Machinery), 586–590. doi: 10.1145/3298689.3347052

- Ferraro, A., Jeon, J. H., Kim, B., Serra, X., and Bogdanov, D. (2020). "Artist biases in collaborative filtering for music recommendation," in *Proceedings of the 37th International Conference on Machine Learning, Vol. 119 of ICML '20*, 1–3. Available online at: <http://hdl.handle.net/10230/45185> (accessed June 15, 2022).
- Ferraro, A., Serra, X., and Bauer, C. (2021a). "Break the loop: gender imbalance in music recommenders," in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, CHIIR '21* (New York, NY: Association for Computing Machinery), 249–254. doi: 10.1145/3406522.3446033
- Ferraro, A., Serra, X., and Bauer, C. (2021b). "What is fair? Exploring the artists' perspective on the fairness of music streaming platforms," in *Human-Computer Interaction-INTERACT 2021: 18th IFIP TC 13 International Conference, Vol. 12933 of Lecture Notes in Computer Science*, eds C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen (Cham: Springer), 562–584. doi: 10.1007/978-3-030-85616-8\_33
- Figueiredo, F., Almeida, J. M., Gonçalves, M. A., and Benevenuto, F. (2014). On the dynamics of social media popularity: a youtube case study. *ACM Trans. Intern. Technol.* 14,24. doi: 10.1145/2665065
- Flexer, A., Dörfler, M., Schlüter, J., and Grill, T. (2018). "Hubness as a case of technical algorithmic bias in music recommendation," in *2018 IEEE International Conference on Data Mining Workshops, ICDMW '18* (New York, NY: Institute of Electrical and Electronics Engineers), 1062–1069. doi: 10.1109/ICDMW.2018.00154
- Folger, R. G., and Cropanzano, R. (1998). *Organizational Justice and Human Resource Management, Vol. 7*. Thousand Oaks, CA: Sage. doi: 10.4135/9781452225777
- Helberger, N., Araujo, T., and de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Comput. Law Secur. Rev.* 39,16. doi: 10.1016/j.clsr.2020.105456
- Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., and Wallach, H. (2019). "Improving fairness in machine learning systems: what do industry practitioners need?," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19* (New York, NY: Association for Computing Machinery), 1–16. doi: 10.1145/3290605.3300830
- Htun, N. N., Lecluse, E., and Verbert, K. (2021). "Perception of fairness in group music recommender systems," in *Proceedings of the 26th International Conference on Intelligent User Interfaces, IUI 2021* (New York, NY: Association for Computing Machinery), 302–306. doi: 10.1145/3397481.3450642
- Hutchinson, B., and Mitchell, M. (2019). "50 years of test (un)fairness: Lessons for machine learning," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT\* '19* (New York, NY: Association for Computing Machinery), 49–58. doi: 10.1145/3287560.3287600
- IFPI (2020). *Global Music Report 2020: The Industry in 2019*. Available online at: [https://www.ifpi.org/wp-content/uploads/2020/07/Global\\_Music\\_Report-the\\_Industry\\_in\\_2019-en.pdf](https://www.ifpi.org/wp-content/uploads/2020/07/Global_Music_Report-the_Industry_in_2019-en.pdf) (accessed June 15, 2022).
- Jannach, D., and Bauer, C. (2020). Escaping the mcnamara fallacy: towards more impactful recommender systems research. *AI Mag.* 41, 79–95. doi: 10.1609/aimag.v41i4.5312
- Jannach, D., Lerche, L., Kamehkhosh, I., and Jugovac, M. (2015). What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Model. User-Adapt. Interact.* 25, 427–491. doi: 10.1007/s11257-015-9165-3
- Kowald, D., Müllner, P., Zangerle, E., Bauer, C., Schedl, M., and Lex, E. (2021). Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Sci.* 10,14. doi: 10.1140/epjds/s13688-021-00268-9
- Kowald, D., Schedl, M., and Lex, E. (2020). "The unfairness of popularity bias in music recommendation: a reproducibility study," in *Advances in Information Retrieval*, eds J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins (Cham: Springer International Publishing), 35–42. doi: 10.1007/978-3-030-45442-5\_5
- Lee, J. H., Pritchard, L., and Hubbles, C. (2019). "Can we listen to it together?: factors influencing reception of music recommendations and post-recommendation behavior," in *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR '19* (Montréal, QC), 663–669.
- Lesota, O., Melchiorre, A., Rekabsaz, N., Brandl, S., Kowald, D., Lex, E., et al. (2021). "Analyzing item popularity bias of music recommender systems: are different genders equally affected?," in *Proceedings of the Fifteenth ACM Conference on Recommender Systems, RecSys '21* (New York, NY: Association for Computing Machinery), 601–606. doi: 10.1145/3460231.3478843
- Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., and Diaz, F. (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* (New York, NY: Association for Computing Machinery), 2243–2251. doi: 10.1145/3269206.3272027
- Mehrotra, R., Xue, N., and Lalmas, M. (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* (New York, NY: Association for Computing Machinery), 3224–3233. doi: 10.1145/3394486.3403374
- Melchiorre, A. B., Rekabsaz, N., Parada-Cabaleiro, E., Brandl, S., Lesota, O., and Schedl, M. (2021). Investigating gender fairness of recommendation algorithms in the music domain. *Inform. Process. Manage.* 58,102666. doi: 10.1016/j.ipm.2021.102666
- Melchiorre, A. B., Zangerle, E., and Schedl, M. (2020). "Personality bias of music recommendation algorithms," in *Proceedings of the Fourteenth ACM Conference on Recommender Systems, RecSys '20* (New York, NY: Association for Computing Machinery), 533–538. doi: 10.1145/3383313.3412223
- Mousavifar, S. M., and Vassileva, J. (2022). "Investigating the efficacy of persuasive strategies on promoting fair recommendations," in *Persuasive Technology, PERSUASIVE 2022*, eds N. Baghaei, J. Vassileva, R. Ali, and K. Oyibo (Cham: Springer International Publishing), 120–133. doi: 10.1007/978-3-030-98438-0\_10
- Neophytou, N., Mitra, B., and Stinson, C. (2022). "Revisiting popularity and demographic biases in recommender evaluation and effectiveness," in *Advances in Information Retrieval, ECIR '22*, eds M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, and V. Setty (Cham: Springer International Publishing), 641–654. doi: 10.1007/978-3-030-99736-6\_43
- Oliveira, R. S., Nóbrega, C., Marinho, L. B., and Andrade, N. (2017). "A multiobjective music recommendation approach for aspect-based diversification," in *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR '17* (Singapore), 414–420. Available online at: <https://archives.ismir.net/ismir2017/paper/000153.pdf> (accessed June 15, 2022).
- Patro, G. K., Biswas, A., Ganguly, N., Gummadi, K. P., and Chakraborty, A. (2020). "Fairrec: two-sided fairness for personalized recommendations in two-sided platforms," in *Proceedings of The Web Conference 2020, WWW '20* (New York, NY: Association for Computing Machinery), 1194–1204. doi: 10.1145/3366423.3380196
- Schedl, M. (2016). "The LFM-1B dataset for music retrieval and recommendation," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, ICMR '16* (New York, NY: Association for Computing Machinery), 103–110. doi: 10.1145/2911996.2912004
- Schedl, M., and Bauer, C. (2017). "Distance- and rank-based music mainstreamness measurement," in *2nd Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems, in Conjunction With 25th International Conference on User Modeling, Adaptation and Personalization (UMAP '17), SOAP 2017* (New York, NY: Association for Computing Machinery), 364–367. doi: 10.1145/3099023.3099098
- Schedl, M., Brandl, S., Lesota, O., Parada-Cabaleiro, E., Penz, D., and Rekabsaz, N. (2022). "LFM-2B: a dataset of enriched music listening events for recommender systems research and fairness analysis," in *ACM SIGIR Conference on Human Information Interaction and Retrieval, CHIIR '22* (New York, NY: Association for Computing Machinery), 337–341. doi: 10.1145/3498366.3505791
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). "Fairness and abstraction in sociotechnical systems," in *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT\* '19* (New York, NY: Association for Computing Machinery), 59–68. doi: 10.1145/3287560.3287598
- Shakespeare, D., Porcaro, L., Gómez, E., and Castillo, C. (2020). "Exploring artist gender bias in music recommendation," in *Proceedings of the Workshops on Recommendation in Complex Scenarios and the Impact of Recommender Systems Co-located With 14th ACM Conference on Recommender Systems (RecSys '20), volume 2697 of ComplexRec-ImpactRS 2020* (New York, NY: Association for



- Computing Machinery), 1–9. Available online at: [http://ceur-ws.org/Vol-2697/paper1\\_impactrs.pdf](http://ceur-ws.org/Vol-2697/paper1_impactrs.pdf) (accessed June 15, 2022).
- Singh, A., and Joachims, T. (2018). “Fairness of exposure in rankings,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18* (New York, NY: Association for Computing Machinery), 2219–2228. doi: 10.1145/3219819.3220088
- Smets, A., Hendrickx, J., and Ballon, P. (2022). We’re in this together: a multi-stakeholder approach for news recommenders. *Digit. J.* 1–19. doi: 10.1080/21670811.2021.2024079
- Sonboli, N., Smith, J. J., Cabral Berenfus, F., Burke, R., and Fiesler, C. (2021). “Fairness and transparency in recommendation: the users’ perspective,” in *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21* (New York, NY: Association for Computing Machinery), 274–279. doi: 10.1145/3450613.3456835
- Stoikov, S., and Wen, H. (2021). “Evaluating music recommendations with binary feedback for multiple stakeholders,” in *Proceedings of the 1st Workshop on Multi-Objective Recommender Systems (MORS 2021) co-located With 15th ACM Conference on Recommender Systems (RecSys 2021), MORS '21*, 1–7. Available online at: <http://ceur-ws.org/Vol-2959/paper9.pdf> (accessed June 15, 2022).
- Tyler, T. R., and Smith, H. J. (1998). “Social justice and social movements,” in *The Handbook of Social Psychology*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (Boston, MA: McGraw-Hill), 595–629.
- Unger, M., Li, P., Cohen, M. C., Brost, B., and Tuzhilin, A. (2021). Deep multi-objective multi-stakeholder music recommendation. *NYU Stern School Bus. Forthcoming*. 1–47. doi: 10.2139/ssrn.3848670
- Wang, Y., and Horvát, E.-Á. (2019). “Gender differences in the global music industry: evidence from musicbrainz and the echo nest,” in *Proceedings of the 13th International AAAI Conference on Web and Social Media, number 01 in ICWSM '19* (Palo Alto, CA: Association for the Advancement of Artificial Intelligence), 517–526. Available online at: <https://ojs.aaai.org/index.php/ICWSM/article/view/3249/3117> (accessed June 15, 2022).
- Wei, T., Feng, F., Chen, J., Wu, Z., Yi, J., and He, X. (2021). “Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21* (New York, NY: Association for Computing Machinery), 1791–1800. doi: 10.1145/3447548.3467289
- Youngs, I. (2019). *Pop Music’s Growing Gender Gap Revealed in the Collaboration Age*. BBC. Available online at: <https://www.bbc.com/news/entertainment-arts-47232677> (accessed June 15, 2022).
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Dinnissen and Bauer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.