It's Not You, It's Me: The Impact of Choice Models and Ranking Strategies on Gender Imbalance in Music Recommendation

Andres Ferraro
andresferraro@acm.org
Pandora-Sirius XM
Oakland, CA, USA

Michael D. Ekstrand
mdekstrand@drexel.edu
Drexel University
Philadelphia, PA, USA

Christine Bauer
christine.bauer@plus.ac.at
Paris Lodron University Salzburg
Salzburg, Austria

ABSTRACT

As recommender systems are prone to various biases, mitigation approaches are needed to ensure that recommendations are fair to various stakeholders. One particular concern in music recommendation is artist gender fairness. Recent work has shown that the gender imbalance in the sector translates to the output of music recommender systems, creating a feedback loop that can reinforce gender biases over time.

In this work, we examine that feedback loop to study whether algorithmic strategies or user behavior are a greater contributor to ongoing improvement (or loss) in fairness as models are repeatedly re-trained on new user feedback data. We simulate user interaction and re-training to investigate the effects of ranking strategies and user choice models on gender fairness metrics. We find re-ranking strategies have a greater effect than user choice models on recommendation fairness over time.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → Collaborative filtering; • **Social and professional topics** → **Gender**;

KEYWORDS

User choice models, re-ranking, artists, music, gender, fairness, bias

ACM Reference Format:
Andres Ferraro, Michael D. Ekstrand, and Christine Bauer. 2024. It's Not You, It's Me: The Impact of Choice Models and Ranking Strategies on Gender Imbalance in Music Recommendation. In *18th ACM Conference on Recommender Systems (RecSys '24), October 14–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3640457.3688163

1 INTRODUCTION

Recommender systems significantly impact users' activities on a wide range of platforms. Streaming platforms have become one of the primary sources of music consumption [22]. Typically, such platforms integrate music recommender systems (MRS) that learn from large-scale user behavior and music features [33] to recommend music (at various levels, including songs, artists, etc.) tailored to a specific user. With the prevalence and impact of automatic recommendation in music (and other domains), it is vital to consider the *fairness* of both the recommendations themselves and their interaction with users and the broader sociotechnical context [8].

In this work, we specifically examine biases related to artist gender. Gender imbalance is a highly topical subject in the music sector [e.g. 20, 34, 37, 39]. From interviews with artists, Ferraro et al. [17] learn that artists care about the gender imbalance in the music industry, a finding partially supported by interviews in Dinnissen and Bauer [5] as well. As artists consider MRS a potential solution to promote content by female artists to reach a gender balance in what users consume [17], Ferraro et al. [16] analyze MRS approaches regarding gender bias and propose bias mitigation strategies to counteract the gender imbalance. In a simulation study, they demonstrate that gradually increasing exposure of underrepresented gender group can interrupt long-term bias amplification. In a field study on Spotify, Epps-Darling et al. [13] found that increasing the fraction of tracks by female artists in algorithmic music recommendations resulted in continuing increase in the prevalence of tracks by female artists in users' "organic" listening (selecting and playing songs without the recommender). Compared to other domains, music has a particularly biased starting point with respect to gender in the underlying data because there is a substantial gender gap: male artists dominate the field, while female and nonbinary artists form a smaller group, though some within this minority group are highly popular [20, 34].

In this work, we build on the findings of Ferraro et al. [16] and Bauer and Ferraro [1], using a similar simulation approach to explore the effects of different post-processing strategies for bias mitigation. In particular, we expand those results to additional post-processing strategies and examine the effects of two different elements of the feedback loop: the **re-ranking strategy** to augment the fairness of the recommendations, and the **user choice model** that encodes how users select results from recommendations. This allows us to examine whether algorithmic strategies or user behavior are a greater contributor to ongoing improvement (or loss) in fairness. We note that this study is purely descriptive—we make no claims about what the balance of artist genders in recommendation should be, but seek to document how that balance is affected by various elements of the sociotechnical feedback loop. While our work focuses specifically on gender imbalance in the music industry, our contribution is relevant to other provider characteristics and content domains, such as books [9].

2 RELATED WORK

There is a wealth of research on biases in recommender systems. Popularity bias has been widely researched for many years [e.g. 2,

RecSys '24, October 14–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0505-2/24/10
https://doi.org/10.1145/3640457.3688163

10, 12, 24]. More recently, research on bias and bias mitigation have been particularly addressed from the perspective of societal fairness: are users or item providers treated fairly—and if not, how can we create fair(er) recommender systems? Biases in recommender system behavior and outputs can arise from many different sources, including the algorithms themselves, corpus data, training data, and users' ongoing interactions with the system [8]. Unfairness arising from both underlying data and recommendation models, and their interplay, has been documented in prior studies [e.g. 9, 16].

In this paper, we are concerned with *fair exposure* or *visibility* for artists across genders. Re-ranking methods are widely adopted to improve such outcome fairness [13, 16, 29, 35, 36, 38], post-processing the outputs of standard recommendation models to improve the fairness of the resulting lists (or, in some cases, grids or other displays). In the context of gender (im)balance, these strategies can be deployed both to improve equity in the recommendation themselves—increasing the exposure that items created by female and gender-minority artists receive—and also to influence future consumption patterns towards gender parity [13].

Most work on fair recommendation is concerned with fairness at a single point in time [e.g. 9]; Epps-Darling et al. [13] looked at the recommender impact on users but did not close the loop for user impact on subsequent recommendations. Real recommender systems, however, are iterative systems with a feedback loop between the system and its users. Simulation is a useful mechanism for studying feedback loops: Mansoury et al. [25] and Jannach et al. [23] use simulations to study popularity bias effects, Zhang et al. [41] find that users' high reliance on recommender systems provides suboptimal performance outcomes in the long run, and Chaney et al. [3] document longitudinal convergence in recommendations between users. While relying on the same simulation approach, Ferraro et al. [15] study exposure biases across artists of different styles, whereas Ferraro et al. [16] and Bauer and Ferraro [1] address gender imbalances in artist exposure. .

Many simulations use very simple models of user response, e.g., assuming that a user would consume the top k items [16]. Recent research on user choice models [19] shows that different choice models can influence the overall choice distribution and performance of the recommenders. Further, different recommender algorithms affect users' choices in different ways [18]. It is therefore necessary to employ richer models of user action to study recommender system feedback loops.

Few works compare the relative effects of recommendation algorithms and user choice models. One example is that of Fabbri et al. [14], who studied exposure inequality in people recommender systems. Their stochastic user choice models did not impact the exposure as much as the recommender algorithm and the underlying social graph structure.

Understanding the relative impact of recommendation techniques and user responses remains a gap in understanding the fairness of recommender systems and developing robust, empirically-grounded tools for ensuring recommendation contributes to a fair and equitable society. In this paper, we build on prior work by two of the authors [1, 16] that used simulation to study the imbalance in gender exposure in MRS; we extend their methods to study and compare the impact of base recommendation models, re-ranking strategies, and user choice models on gender balance in an iterative setting where recommendation models are repeatedly re-trained in response to user feedback.

3 METHODS

We build our experiment on a subset of the *LFM-2b* dataset [32]. We use two base recommendation models, *IALS* (implicit-feedback matrix factorization with alternating least squares [21]) and *BPR* (matrix factorization trained with pairwise rank loss [30]), to generate initial recommendation lists and reprocess these lists through various post-processing bias mitigation strategies. We ran the experiments with LensKit [6], using LensKit's *IALS* implementation and a PyTorch *BPR* implementation for LensKit. We tuned the hyperparameters for both base models using random search to optimize *MRR* over 1000-item rankings on a test set.

We first describe the mitigation strategies we study, then remaining details of the data, metrics, and experimental simulation.[1]

3.1 Approaches

We choose *IALS* and *BPR* as the basis approaches for our analysis because these are well-known algorithms for collaborative filtering. Each mitigation strategy post-processes the ranking produced by *IALS* or *BPR* to improve the gender fairness of the final ranking. We test three strategies for recommendation post-processing:

- ***MoveUp*** Move the first item by a female artist to the first rank.
- **λ5, λ7** Penalize items by male artists by moving each of them λ positions downward in the ranking [16]; based on the original paper, we use $\lambda = 5$ and $\lambda = 7$.
- ***FAIR*** The FA*IR ranking algorithm from Zehlike et al. [40]: For each position, select the highest-scored item in the original ranking that will not cause the protected group (non-male artists) to be statistically significantly underrepresented concerning the target proportion (which we set to 50%).

IALS and *BPR* without any adaptations (*None*) serve as baselines.

3.2 Dataset

For our exploratory analysis, we rely on a subset of interactions between 2013–2020 in the *LFM-2b* dataset [32] enriched with artist gender information collected from MusicBrainz.org (*MB*)[2]. We limit the data to only 'solo' artists—where the artist is a person—for which *MB* reports the gender identities. We consider the categories female, male, and nonbinary[3].

The final data contains 436, 789 artists, of which 93, 316 are categorized as female, 342, 523 as male, and 950 as nonbinary. Following the procedure of Ferraro et al. [16], we use a 15-core of user-artist interactions, retaining 78, 021 users and 187, 471 artists, of whom 21.675% represent female artists. For hyperparameter tuning, we split data by time to get 90% training and 10% test data, discarding test users with no training ratings.

3.3 Metrics

We use several metrics to understand the system's behavior from different perspectives:

[1] Code available at https://doi.org/10.5281/zenodo.13315571.
[2] https://musicbrainz.org
[3] Nonbinary refers to a broad spectrum of different individual identities. From *MB*, we merged the categories nonbinary and other.

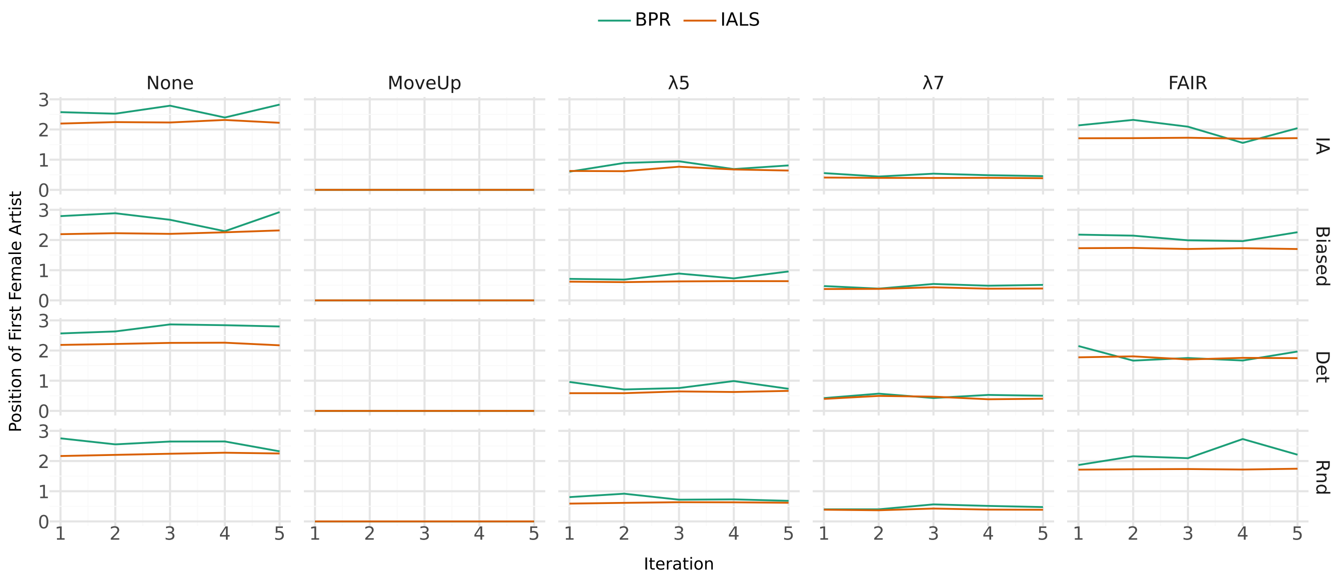

Figure 1: Average position of the first female artist (*PFA*) over iterations. Position 0 is the highest rank.

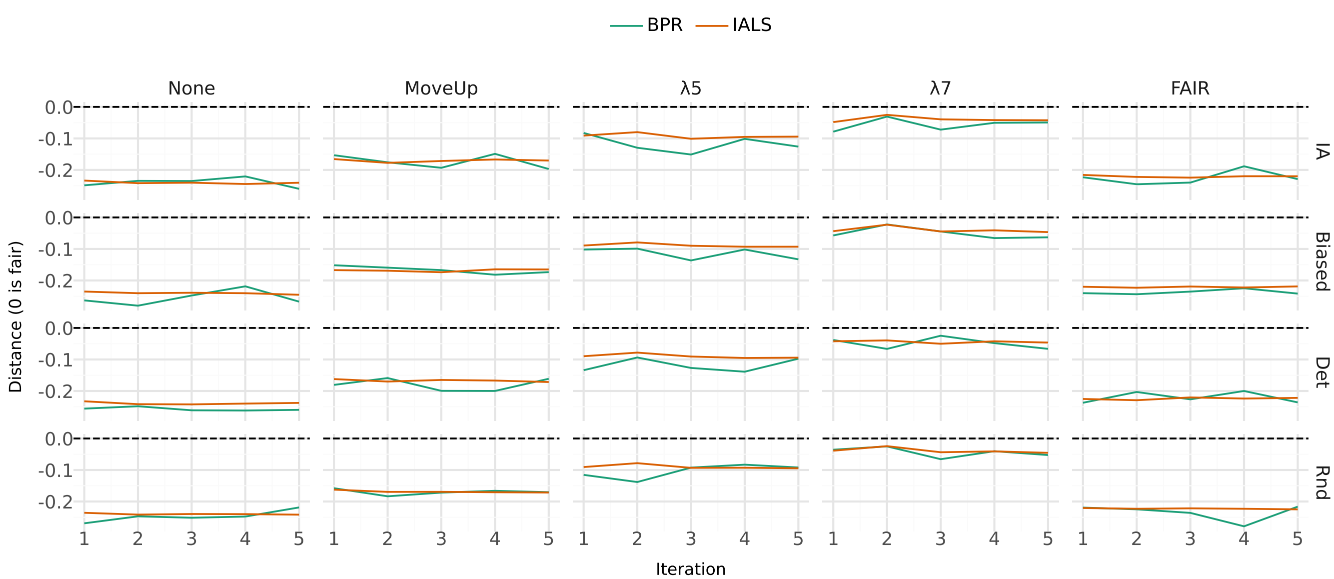

Figure 2: *AWRF* over iterations. 0 is fair; negative values mean that unprotected-group (male) artists are overexposed.

- *First-position exposure.* We focus on the position in the recommendation rankings because users interact more frequently with only the top-ranked items (i.e., position bias) [4]. To this end—in line with Ferraro et al. [16], we average for each user the *position of the first female artist* (*PFA*) in the recommendation ranking, with the first position as 0.
- *Overall exposure* with *AWRF* (Attention-Weighted Rank Fairness [28, 31]). This uses rank-discounting (as in metrics like *nDCG* and *RBP*; we followed TREC in using *nDCG*'s logarithmic discounting [11]) to estimate the exposure value of each rank position, and measuring the fairness of exposure provided to each group. As with *FAIR*, we treat male artists as unprotected, and measure fairness by the difference between the exposure distribution in the ranking and a target of 50%, and average over all recommendation lists.
- *Diversity.* We use the *Gini index* (*Gini@k*) to measure how concentrated the recommendations are on a few artists both

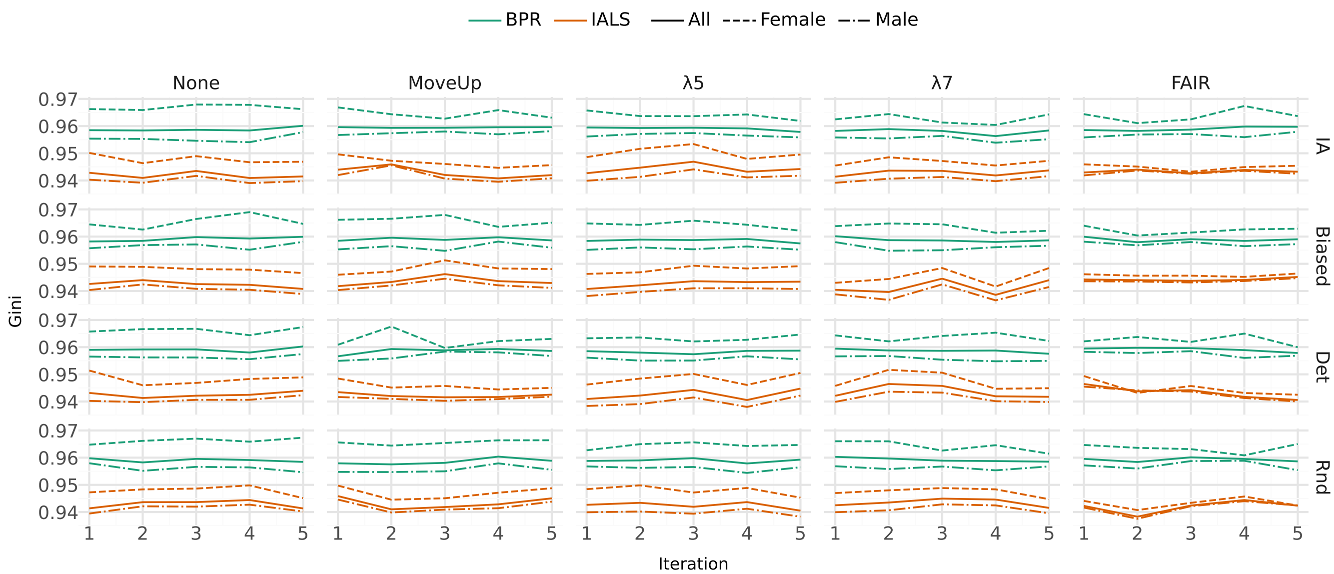

Figure 3: Gini coefficients (*Gini*@10) over iterations, both overall and for male and female artists. 1 **is maximum inequality.**

overall and disaggregated by gender. $Gini@k = 1$ indicates that all recommendations go to the same artist, while 0 means all artists are equally recommended.

3.4 User Choice Models

In the simulation, we consider four different user choice models (each implemented for $N = 10$):

Deterministic (*Det*) User listens to all N recommended items.

Random (*Rnd*) For each item in a user's top N recommended items, the user listens with probability 0.5.

InspectionAbandon (*IA*) A probabilistic cascade model [27]: for each item in the top-N recommendations, with probability 0.5 the user listens to that item. After listening or ignoring the item, they stop listening entirely with probability 0.3; otherwise, they continue to the next item. The *IA* model is based on cascade models from the existing evaluation literature [27], implementing that users have a higher probability of consuming higher-ranked items than lower-ranked ones.

Biased Variant of *IA* in which users are biased so that they are 10% more likely to listen to male artists than others.

3.5 Simulation

We simulate a user-recommender feedback loop to study the long-term effect of these mitigation strategies and user choice models. For each user, we use the choice model to select artists from the system's recommendations and record additional listening events for the selected artists. We then retrain the model on this augmented data and compute recommendations for the next iteration. We repeat this procedure for 5 iterations, as the effects anchored after 5 iterations in previous works [16].

4 RESULTS

In this section, we present the results concerning the *position of the first female artist* (*PFA*), *AWRF*, and the *Gini coefficients*. We compare the re-ranking approaches (in the columns of the facet plots in the figures) and the user choice models (in the rows of the facet plots). While it is important to include the full gender spectrum in the data, model, and simulation, our data does not contain sufficient nonbinary artists to draw scientific conclusions from their inclusion in the analysis, so we report results only on *male* and *female* artists.

Fig. 1 clearly indicates that the re-ranking strategy determines the position of the first female artist, whereas the user choice model does not impact the ranking position of the first female artist. *IALS* is more stable than *BPR*, showing almost no variation across iterations or choice models. As *MoveUp* always puts a female artist on the very first position, it optimizes this metric by design. $\lambda 7$—followed by $\lambda 5$—ranks the first female artist only on a slightly lower rank. *FAIR* ranks the first female artist only slightly higher than the baselines without re-ranking (*None*).

AWRF is mostly consistent with *PFA*. Fig. 2 indicates that the re-ranking strategies have more impact than user choice models on *AWRF* (there is more variation between columns than rows), with little change over time. Further, we observe that *FAIR* contributes least to improving *AWRF* compared to the baselines; moving only the first female artist (*MoveUp*) is more effective at overall exposure fairness even though it only adjusts the position of a single item. The λ-re-rankers were the most effective, with $\lambda 7$ improving exposure fairness the most. As with *PFA*, *IALS*-based recommendations had more stable gender exposure balance.

Fig. 3 shows diversity as measured by *Gini*@10. *IALS* is more equitable than *BPR* across all re-ranking strategies and choice models. Further, comparing $Gini@10_{female}$ and $Gini@10_{male}$ shows more equitable results for male than female artists: not only are female artists under-recommended (see *AWRF*), the exposure that does go to female artists is more concentrated on a smaller fraction of those artists. With *FAIR*, the gender gap in inequity closes, particularly with the *IALS* base model.

5 DISCUSSION AND CONCLUSION

Our results show that re-ranking strategies have a greater effect than user choice models on the provider fairness of recommendations as the model is retrained with new user interactions over time: when it comes to longitudinal fairness, 'it's not you' (the users), 'it's me' (the recommender system). This effect is consistent across multiple metrics and underlying recommendation models. That is, the model of user choice and response to recommendation had little effect on the fairness metrics we considered. Therefore, re-ranking strategies may be a useful tool for intervening in a biased world and addressing 'societal imbalance' [26], as the algorithms have a stronger impact than the users' behavior in choosing items. The base recommendation models we considered did exhibit different behavior, with *IALS* delivering more stable results while *BPR* showed greater variation in fairness and diversity metrics, although without clear trends. While the re-ranking strategies can break the original feedback loop, *IALS* seems to anchor a 'new' feedback loop within the 5 iterations observed in our study.

It is not clear to what extent *BPR*'s volatility is a response to user data vs. noise and sensitivity to randomness in training. Isolating these effects would require repeated re-runs of the simulation, which is computationally expensive; developing sample-efficient ways of identifying sources of noise [7] is an important future research direction for recommender system simulation.

There are several important directions to extend and improve this work. For one, while we considered nonbinary gender identities, the small number of nonbinary artists makes meaningful analysis with current methods difficult. We plan future work to explore additional re-ranking strategies and choice models (including more heavily biased models), additional data sets and domains, and a broader range of recommendation strategies (including content-based and hybrid models). It is also vital in the future to examine intersectional impacts on artists who belong to multiple marginalized groups.

ACKNOWLEDGMENTS

This publication was supported by the Excellence in Digital Sciences and Interdisciplinary Technologies (EXDIGIT) project, funded by Land Salzburg under grant number 20204-WISS/263/6-6022, and by the U.S. National Science Foundation under grant number 24-15042.

REFERENCES

[1] Christine Bauer and Andres Ferraro. 2023. Strategies for Mitigating Artist Gender Bias in Music Recommendation: a Simulation Study. In *Proceedings of the Music Recommender Systems Workshop* (Singapore) *(MuRS 2023)*. Zenodo, 5 pages. https://doi.org/10.5281/zenodo.8372477

[2] Òscar Celma and Pedro Cano. 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* (Las Vegas, NV, USA) *(NETFLIX '08)*. Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/1722149.1722154

[3] Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. 2018. How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, BC, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 224–232. https://doi.org/10.1145/3240323.3240370

[4] Andrew Collins, Dominika Tkaczyk, Akiko Aizawa, and Joeran Beel. 2018. Position Bias in Recommender Systems for Digital Libraries. In *Transforming Digital Worlds* (Sheffield, UK) *(iConference 2018)*. Springer, Cham, Switzerland, 335–344. https://doi.org/10.1007/978-3-319-78105-1_37

[5] Karlijn Dinnissen and Christine Bauer. 2023. Amplifying Artists' Voices: Item Provider Perspectives on Influence and Fairness of Music Streaming Platforms. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (Limassol, Cyprus) *(UMAP '23)*. Association for Computing Machinery, New York, NY, USA, 238–249. https://doi.org/10.1145/3565472.3592960

[6] Michael D. Ekstrand. 2020. LensKit for Python: Next-Generation Software for Recommender Systems Experiments. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 2999–3006. https://doi.org/10.1145/3340531.3412778

[7] Michael D. Ekstrand, Ben Carterette, and Fernando Diaz. 2024. Distributionally-Informed Recommender System Evaluation. *ACM Transactions on Recommender Systems* 2, 1, Article 6 (March 2024), 27 pages. https://doi.org/10.1145/3613455

[8] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Information Retrieval* 16, 1–2 (2022), 196 pages. https://doi.org/10.1561/1500000079

[9] Michael D. Ekstrand and Daniel Kluver. 2021. Exploring Author Gender in Book Rating and Recommendation. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 377–420. https://doi.org/10.1007/s11257-020-09284-2

[10] Michael D. Ekstrand and Vaibhav Mahant. 2017. Sturgeon and the Cool Kids: Problems with Top-N Recommender Evaluation. In *Proceedings of the 30th Florida Artificial Intelligence Research Society Conference* (Marco Island, FL, USA) *(FLAIRS 2017)*. AAAI Press, Washington, DC, USA, 639–644. https://aaai.org/papers/639-flairs-2017-15534/

[11] Michael D Ekstrand, Graham McDonald, Amifa Raj, and Isaac Johnson. 2022. Overview of the TREC 2021 Fair Ranking Track. In *The Thirtieth Text REtrieval Conference (TREC 2021) Proceedings*. https://trec.nist.gov/pubs/trec30/papers/Overview-F.pdf

[12] Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Sorush Saghari, Shiva Parsa Rad, and Marko Tkalčič. 2021. Investigating the Impact of Recommender Systems on User-based and Item-based Popularity Bias. *Information Processing & Management* 58, 5, Article 102655 (Sept. 2021), 15 pages. https://doi.org/10.1016/j.ipm.2021.102655

[13] Avriel Epps-Darling, Romain Takeo Bouyer, and Henriette Cramer. 2020. Artist Gender Representation in Music Streaming. In *Proceedings of the 21st International Society for Music Information Retrieval Conference* (Montréal, Canada) *(ISMIR 2020)*. ISMIR, 248–254. https://doi.org/10.5281/zenodo.4245416

[14] Francesco Fabbri, Maria Luisa Croci, Francesco Bonchi, and Carlos Castillo. 2022. Exposure Inequality in People Recommender Systems: The Long-Term Effects. *Proceedings of the International AAAI Conference on Web and Social Media* 16, 1 (May 2022), 194–204. https://doi.org/10.1609/icwsm.v16i1.19284

[15] Andres Ferraro, Dmitry Bogdanov, Xavier Serra, and Jason Yoon. 2019. Artist and Style Exposure Bias in Collaborative Filtering Based Music Recommendations. In *Proceedings of the 1st Workshop on Designing Human-Centric Music Information Research Systems* (Delft, The Netherlands) *(wsHCMIR '19)*. 8–10. https://doi.org/10.48550/arXiv.1911.04827

[16] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the Loop: Gender Imbalance in Music Recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra, ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 249–254. https://doi.org/10.1145/3406522.3446033

[17] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. 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* (Bari, Italy) *(INTERACT '21, Vol. 12933)*. Springer, Cham, Switzerland, 562–584. https://doi.org/10.1007/978-3-030-85616-8_33

[18] Naieme Hazrati and Francesco Ricci. 2022. Recommender Systems Effect on the Evolution of Users' Choices Distribution. *Information Processing & Management* 59, 1, Article 102766 (2022), 18 pages. https://doi.org/10.1016/j.ipm.2021.102766

[19] Naieme Hazrati and Francesco Ricci. 2024. Choice Models and Recommender Systems Effects on Users' Choices. *User Modeling and User-Adapted Interaction* 34, 1 (2024), 109–145. https://doi.org/10.1007/s11257-023-09366-x

[20] Karla Hernandez, Stacy L. Smith, Marc Choueiti, and Katherine Pieper. 2022. *Inclusion in the Recording Studio?: Gender and Race/Ethnicity of Artists, Songwriters & Producers across 1,000 Popular Songs from 2012–2021*. Technical Report. Annenberg Inclusion Initiative. https://assets.uscannenberg.org/docs/aii-inclusion-recording-studio-20220331.pdf

[21] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *2008 Eighth IEEE International Conference on Data Mining* (Pisa, Italy) *(ICDM 2008)*. IEEE, New York, NY, USA, 263–272. https://doi.org/10.1109/ICDM.2008.22

[22] IFPI. 2024. *IFPI Global Music Report 2024: State of the Industry.* Technical Report. IFPI, London, UK. https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/IFPI_GMR_2024_State_of_the_Industry_db92a1c9c1.pdf

[23] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What Recommenders Recommend: an Analysis of Recommendation Biases and Possible Countermeasures. *User Modeling and User-Adapted Interaction* 25, 5 (2015), 427–491. https://doi.org/10.1007/s11257-015-9165-3

[24] Dominik Kowald and Emanuel Lacic. 2022. Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems. In *Advances in Bias and Fairness in Information Retrieval (BIAS 2022, Vol. 1610)*. Springer, Cham, Switzerland, 1–11. https://doi.org/10.1007/978-3-031-09316-6_1

[25] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback Loop and Bias Amplification in Recommender Systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 2145–2148. https://doi.org/10.1145/3340531.3412152

[26] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application* 8 (Nov. 2021), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

[27] Amifa Raj and Michael Ekstrand. 2023. Unified Browsing Models for Linear and Grid Layouts. *CoRR* arXiv:2310.12524 (2023), 7 pages. https://doi.org/10.48550/arXiv.2310.12524

[28] Amifa Raj and Michael D Ekstrand. 2022. Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 726–736. https://doi.org/10.1145/3477495.3532018

[29] Amifa Raj and Michael D. Ekstrand. 2024. Towards Optimizing Ranking in Grid-Layout for Provider-Side Fairness. In *Advances in Information Retrieval* (Glasgow, UK) *(ECIR 2024, Vol. 14612)*. Springer, Cham, Switzerland, 90–105. https://doi.org/10.1007/978-3-031-56069-9_7

[30] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (Montreal, Quebec, Canada) *(UAI '09)*. AUAI Press, Arlington, VA, USA, 452–461. https://doi.org/10.5555/1795114.1795167

[31] Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. 2019. Quantifying the Impact of User Attentionon Fair Group Representation in Ranked Lists. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 553–562. https://doi.org/10.1145/3308560.3317595

[32] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekabsaz. 2022. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 337–341. https://doi.org/10.1145/3498366.3505791

[33] Markus Schedl, Peter Knees, Brian McFee, and Dmitry Bogdanov. 2022. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, USA, 927–971. https://doi.org/10.1007/978-1-0716-2197-4_24

[34] Stacy L. Smith, Katherine Pieper, Karla Hernandez, and Sam Wheeler. 2024. *Inclusion in the Recording Studio?: Gender and Race/Ethnicity of Artists, Songwriters & Producers across 1,200 Popular Songs from 2012–2023.* Technical Report. Annenberg Inclusion Initiative. https://assets.uscannenberg.org/docs/aii-inclusion-recording-studio-20240130.pdf

[35] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-aspect Fairness through Personalized Re-ranking. In *UMAP '20*. Association for Computing Machinery, New York, NY, USA, 239–247. https://doi.org/10.1145/3340631.3394846

[36] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-aspect Fairness through Personalized Re-ranking. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 239–247. https://doi.org/10.1145/3340631.3394846

[37] Yixue Wang and Emőke-Ágnes Horvát. 2019. Gender Differences in the Global Music Industry: Evidence from MusicBrainz and The Echo Nest. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 01 (July 2019), 517–526. https://doi.org/10.1609/icwsm.v13i01.3249

[38] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A Survey on the Fairness of Recommender Systems. *ACM Transactions on Information Systems* 41, 3, Article 52 (Feb. 2023), 43 pages. https://doi.org/10.1145/3547333

[39] Jada Watson. 2020. Programming Inequality: Gender Representation on Canadian Country Radio (2005–2019). In *Proceedings of the 21st International Society for Music Information Retrieval Conference* (Montréal, Canada) *(ISMIR 2020)*. ISMIR, 392–399. https://doi.org/10.5281/zenodo.4245452

[40] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A Fair Top-k Ranking Algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Singapore) *(CIKM '17)*. Association for Computing Machinery, New York, NY, USA, 1569–1578. https://doi.org/10.1145/3132847.3132938

[41] Jingjng Zhang, Gediminas Adomavicius, Alok Gupta, and Wolfgang Ketter. 2020. Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework. *Information Systems Research* 31, 1 (2020), 76–101. https://doi.org/10.1287/isre.2019.0876