

あなたではありません、私です：音楽推薦におけるジェンダー不均衡に対する選択モデルとランキング戦略の影響

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

推薦システムは様々なバイアスに陥りやすいため、推薦が様々な利害関係者にとって公平であることを保証するために、緩和アプローチが必要である。音楽推薦における特に懸念されるのは、アーティストのジェンダーの公平性である。最近の研究では、この分野におけるジェンダーの不均衡が音楽推薦システムのアウトプットにつながり、時間の経過とともにジェンダーの偏りを強化するフィードバックループが形成されることが示されている。本研究では、新しいユーザフィードバックデータに対してモデルを繰り返し再トレーニングすることで、アルゴリズム戦略やユーザ行動が、継続的な公平性の向上(または損失)に大きく寄与するかどうかを研究するために、そのフィードバックループを検証する。ランキング戦略とユーザー選択モデルがジェンダー公平性メトリクスに与える影響を調査するために、ユーザーとの対話と再トレーニングをシミュレートする。我々は、リランキング戦略が、ユーザー選択モデルよりも、時間の経過とともに推薦の公平性に大きな影響を与えることを発見した。

CCS CONCEPTS

情報システム → 推薦システム; 人間中心コンピューティング → 協調フィルタリング; 社会的・職業的トピック → ジェンダー;

KEYWORDS

ユーザー選択モデル、再ランキング、アーティスト、音楽、ジェンダー、公平性、バイアス

ACMリファレンス形式:

アンドレス・フェラーロ、マイケル・D・エクストランド、クリスティン・バウアー。2024。あなたではありません、私です：音楽推薦におけるジェンダー不均衡に対する選択モデルとランキング戦略の影響。第18回ACM推薦システム会議(RecSys '24)、2024年10月14日~18日、バリ、イタリア。ACM, ニューヨーク, 米国, 6ページ。https://doi.org/10.1145/3640457.3688163

1 INTRODUCTION

推薦システムは、様々なプラットフォームにおけるユーザの活動に大きな影響を与える。ストリーミング・プラットフォームは、音楽消費の主要な供給源の一つとなっている[22]。一般的に、このようなプラットフォームは、大規模なユーザー行動と音楽の特徴から学習する音楽推薦システム(MRS)[33]を統合し、特定のユーザーに合わせた音楽(曲、アーティストなど様々なレベル)を推薦する。



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

RecSys '24, October 14–18, 2024, Bari, Italy
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0505-2/24/10
https://doi.org/10.1145/3640457.3688163

音楽(および他の領域)における自動推薦の普及と影響に伴い、推薦自体とユーザーとのインタラクション、およびより広範な社会技術的文脈の両方の公平性を考慮することが不可欠である[8]。

本研究では、特にアーティストの性別に関連するバイアスを検証する。ジェンダーの不均衡は、音楽分野では非常に話題性の高いテーマである[20, 34, 37, 39など]。アーティストとのインタビューから、Ferraroら[17]は、アーティストが音楽業界におけるジェンダーの不均衡を気にしていることを学び、この発見はDinnissen and Bauer [5]のインタビューでも部分的に支持されている。アーティストは、MRSを女性アーティストがユーザーが消費するジェンダーバランスに到達するためのコンテンツを促進する潜在的なソリューションと考えるため[17]、Ferraroら[16]は、ジェンダーバイアスに関するMRSアプローチを分析し、ジェンダー不均衡に対抗するためのバイアス緩和戦略を提案している。シミュレーション研究において、代表的でないジェンダー・グループへの曝露を徐々に増加させることで、長期的なバイアスの増幅を中断できることを実証した。Spotifyに関するフィールド研究において、Epps-Darlingら[13]は、アルゴリズムによる音楽推薦において、女性アーティストによる楽曲の割合を増やすと、ユーザーの「オーガニック」リスニング(推薦者なしで曲を選択し、再生する)において、女性アーティストによる楽曲の普及率が継続的に増加することを発見した。他の領域と比較して、音楽は、実質的なジェンダーギャップがあるため、基礎となるデータにおいてジェンダーに関して特に偏った出発点を持っている:男性アーティストがフィールドを支配し、女性アーティストとノンバイナリーアーティストはより小さなグループを形成するが、この少数派グループの中には、非常に人気のある者もいる[20, 34]。

本研究では、Ferraroら[16]とBauer and Ferraro[1]の知見を基に、同様のシミュレーションアプローチを用いて、バイアス緩和のための異なる後処理戦略の効果を探索する。特に、これらの結果を追加の後処理戦略に拡張し、フィードバックループの2つの異なる要素の効果を検証する:推薦の公平性を増強する再順位付け戦略と、ユーザが推薦から結果を選択する方法をエンコードするユーザー選択モデルである。これにより、アルゴリズム戦略やユーザー行動が、公平性の継続的な改善(または損失)に大きく寄与しているかどうかを検証することができる。本研究は純粋に記述的なものであり、推薦におけるアーティストのジェンダーのバランスはどうあるべきかについて主張するものではなく、そのバランスが社会技術的フィードバックループの様々な要素によってどのように影響されるかを文書化しようとするものであることに留意されたい。我々の研究は、音楽業界におけるジェンダーの不均衡に特に焦点を当てているが、我々の貢献は、書籍[9]のような他のプロバイダーの特性やコンテンツドメインに関連している。

2 関連研究

推薦システムにおけるバイアスに関する研究は豊富にある。人気度バイアスは長年にわたって広く研究されてきた[例2、



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 Dinissen 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,



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

RecSys '24, October 14–18, 2024, Bari, Italy
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0505-2/24/10
<https://doi.org/10.1145/3640457.3688163>

10, 12, 24]. さらに最近では、バイアスとバイアスの軽減に関する研究が、社会的公正の観点から特に取り上げられている。ユーザーやアイテムプロバイダーは公平に扱われているのか、そうでない場合、どのように公平なレコメンダーシステムを作ることができるのか。推薦システムの動作と出力におけるバイアスは、アルゴリズム自体、コーパスデータ、学習データ、ユーザーとシステムとの継続的な相互作用など、多くの異なるソースから発生する可能性がある[8]。基礎となるデータと推薦モデルの両方から生じる不公平さ、およびそれらの相互作用は、先行研究[9, 16]で文書化されている。本稿では、男女を問わず、アーティストの公正な露出や知名度に関心がある。このような結果の公平性を向上させるために、再順位付け手法が広く採用されており[13, 16, 29, 35, 36, 38]、標準的な推薦モデルの出力を後処理して、結果として得られるリスト(場合によってはグリッドや他のディスプレイ)の公平性を向上させる。ジェンダー(不)バランスの文脈では、これらの戦略は、推薦自体の公平性を向上させ、女性やジェンダー・マイノリティのアーティストが作成するアイテムが受ける露出を増加させ、またジェンダー・パリティに向けた将来の消費パターンに影響を与えるために展開することができる[13]。

公正な推薦に関するほとんどの研究は、ある時点における公平性に関するものである[例えば9]。Epps-Darlingら[13]は、推薦者がユーザーに与える影響を調べたが、その後の推薦に対するユーザーの影響についてはループを閉じることはしなかった。しかし、実際のレコメンダーシステムは、システムとそのユーザーとの間にフィードバックループを持つ反復システムである。シミュレーションはフィードバックループを研究するのに有用なメカニズムである：Mansouryら[25]とJannachら[23]は人気バイアス効果を研究するためにシミュレーションを使用し、Zhangら[41]はユーザーが推薦システムに大きく依存することは、長期的には最適とは言えないパフォーマンス結果をもたらすことを発見し、Chaneyら[3]はユーザー間の推薦における縦断的な収束を文書化している。同じシミュレーション手法に依存しながら、Ferraroら[15]は異なるスタイルのアーティスト間の露出バイアスを研究し、Ferraroら[16]とBauer and Ferraro [1]はアーティスト露出のジェンダー不均衡に対処している。

多くのシミュレーションでは、例えば、ユーザが上位 k 個のアイテムを消費すると仮定した場合など、非常に単純なユーザー応答モデルを使用している[16]。ユーザー選択モデルに関する最近の研究[19]は、異なる選択モデルがレコメンダーの全体的な選択分布と性能に影響を与える可能性があることを示している。さらに、推薦アルゴリズムの違いは、ユーザの選択に異なる影響を与える[18]。したがって、推薦システムのフィードバックループを研究するために、より豊富なユーザーアクションのモデルを採用する必要がある。推薦アルゴリズムとユーザ選択モデルの相対的な効果を比較した研究はほとんどない。その一例が、人物推薦システムにおける露出の不平等を研究したFabbriら[14]のものである。確率的ユーザ選択モデルは、推薦アルゴリズムやその基礎となるソーシャルグラフ構造ほど、露出に影響を与えなかった。推薦技術とユーザーの反応の相対的な影響を理解することは、推薦システムの公平性を理解し、推薦が公平で公平な社会に貢献することを保証するための、強固で経験に基づいたツールを開発する上で、依然としてギャップである。本論文では、MRSにおけるジェンダー暴露の不均衡を研究するためにシミュレーションを使用した2人の著者[1, 16]による先行研究を基に、彼らの方法を拡張して、推薦モデルがユーザーのフィードバックに応じて繰り返し再トレーニングされる反復設定において、

ベース推薦モデル、再ランク戦略、ユーザー選択モデルがジェンダーバランスに与える影響を研究し比較する。

3 METHODS

LFM-2bデータセット[32]のサブセットで実験を構築する。我々は、IALS(交互最小二乗法による暗黙フィードバック行列分解[21])とBPR(ペアワイズランクロス[30]で学習した行列分解)という2つの基本推薦モデルを用いて、初期推薦リストを生成し、様々な後処理バイアス緩和戦略によってこれらのリストを再処理する。Lens KitのIALS実装とLensKitのPyTorch BPR実装を用いて、LensKit [6]で実験を行った。テストセットで1000項目のランキングに対してMRRを最適化するために、ランダムサーチを用いて両ベースモデルのハイパーパラメータをチューニングした。

まず、我々が研究している緩和戦略について説明し、次にデータ、測定基準、実験シミュレーションの詳細を述べる¹。

3.1 Approaches

IALSとBPRは協調フィルタリングのアルゴリズムとしてよく知られているため、分析の基礎となるアプローチとして選択した。各緩和戦略は、最終的なランキングのジェンダー公平性を向上させるために、IALSまたはBPRによって生成されたランキングを後処理する。推薦の後処理として、3つの戦略をテストする：

MoveUp 女性アーティストによる最初のアイテムを1位に移動させる。 $\lambda 5$, $\lambda 7$ 男性アーティストによるアイテムのペナルティは、それぞれ λ の位置をランキングの下方に移動させることで行う[16]。原著論文に基づき、 $\lambda=5$, $\lambda=7$ を使用する。
FAIR Zehlkeら[40]のFA*IRランキングアルゴリズム：各ポジションについて、保護グループ(非男性アーティスト)がターゲット比率(50%に設定)に関して統計的に有意に過小評価されないような、元のランキングで最もスコアの高いアイテムを選択する。

IALS 適応のないIALSとBPR(なし)がベースラインとなる。

3.2 Dataset

探索的分析のために、MusicBrainz.org (MB)² から収集したアーティストの性別情報を強化したLFM-2bデータセット[32]の2013年から2020年の間の相互作用のサブセットに依存する。データを「ソロ」アーティスト(アーティストが人物)のみに限定し、MBがジェンダー・アイデンティティを報告する。女性、男性、ノンバイナリ³のカテゴリーを考える。最終的なデータには436,789人のアーティストが含まれ、そのうち93,316人が女性、342,523人が男性、950人がノンバイナリーに分類される。Ferraroら[16]の手順に従い、ユーザーとアーティストのインタラクションを15コア使用し、78,021人のユーザーと187,471人のアーティストを保持し、そのうち21.675%が女性アーティストである。ハイパーパラメータのチューニングでは、90%のトレーニングデータと10%のテストデータを得るために、時間を分割し、トレーニング評価のないテストユーザーを破棄する。

3.3 Metrics

我々は、システムの動作を異なる視点から理解するために、いくつかのメトリクスを使用する：

コードは<https://doi.org/10.5281/zenodo.13315571>に

ある。²

非二元論とは、個人のアイデンティティの広範なスペクトルを指す。MBから、非二値とその他のカテゴリーを統合した。

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.

$\lambda 5, \lambda 7$ 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 nonbinary³.

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:

¹Code available at <https://doi.org/10.5281/zenodo.13315571>.

²<https://musicbrainz.org>

³Nonbinary refers to a broad spectrum of different individual identities. From *MB*, we merged the categories nonbinary and other.

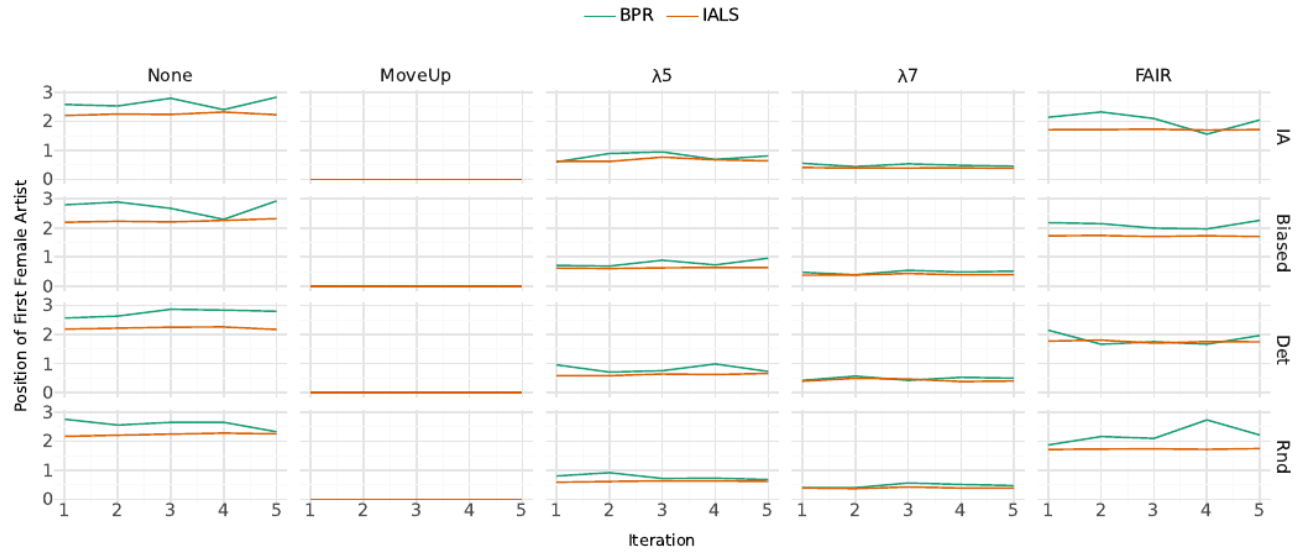


図1:反復における最初の女性アーティスト(PFA)の平均位置。位置 0 が最高ランクである。

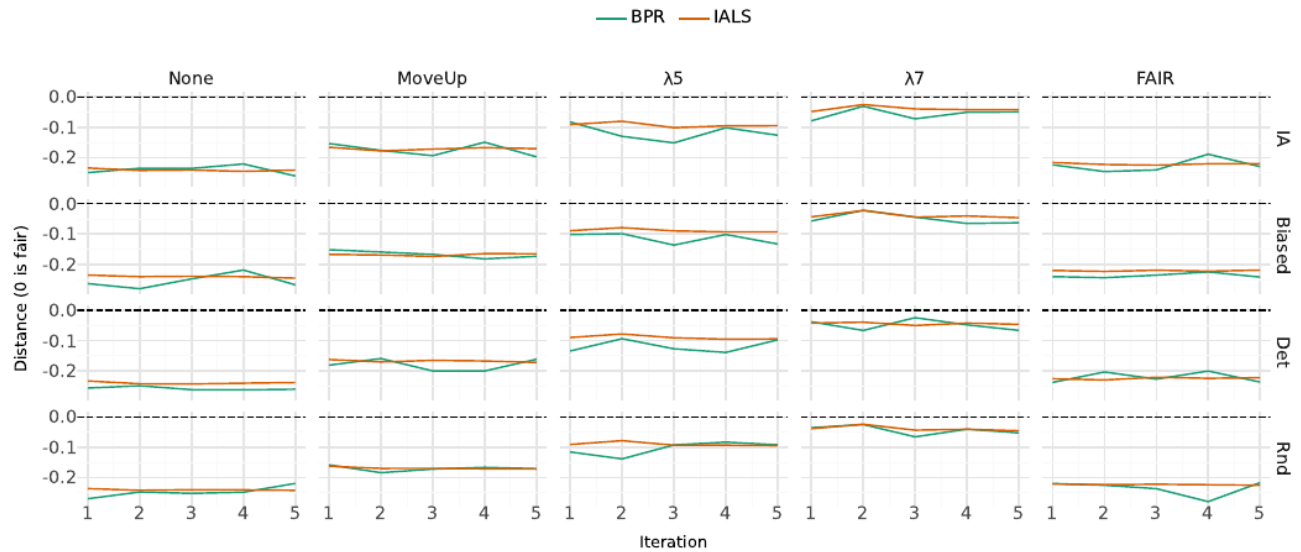


図2:反復回数に対するAW RF。0は公正であり、負の値は保護されていないグループ(男性)のアーティストが過剰に露出していることを意味する。

- 第一位露出。我々は、ユーザが上位にランクされたアイテムのみとより頻繁に対話するため、推薦ランキングにおける位置に注目する(すなわち、位置の偏り)[4]。このため、Ferraroら[16]と同様に、推薦ランキングにおける最初の女性アーティスト(PFA)の位置を各ユーザーについて平均し、最初の位置を0とする。
- AW RF(Attention-Weighted Rank Fairness [28, 31])による全体的な露出。

これは、ランク割引(nDCGやRBPのような指標;我々は、nDCGの対数割引[11]を使用する際にTRECに従った)を使用して、各ランク位置の露出値を推定し、各グループに提供される露出の公平性を測定する。FAIRと同様に、男性アーティストを無防備として扱い、ランキングの露出分布と50%の目標との差で公平性を測定し、すべての推薦リストの平均をとる。

- 多様性。ジニ指数(Gini@k)を用いて、推薦文が全体的に、また性別に細分化された少数のアーティストにどれだけ集中しているかを測定する。

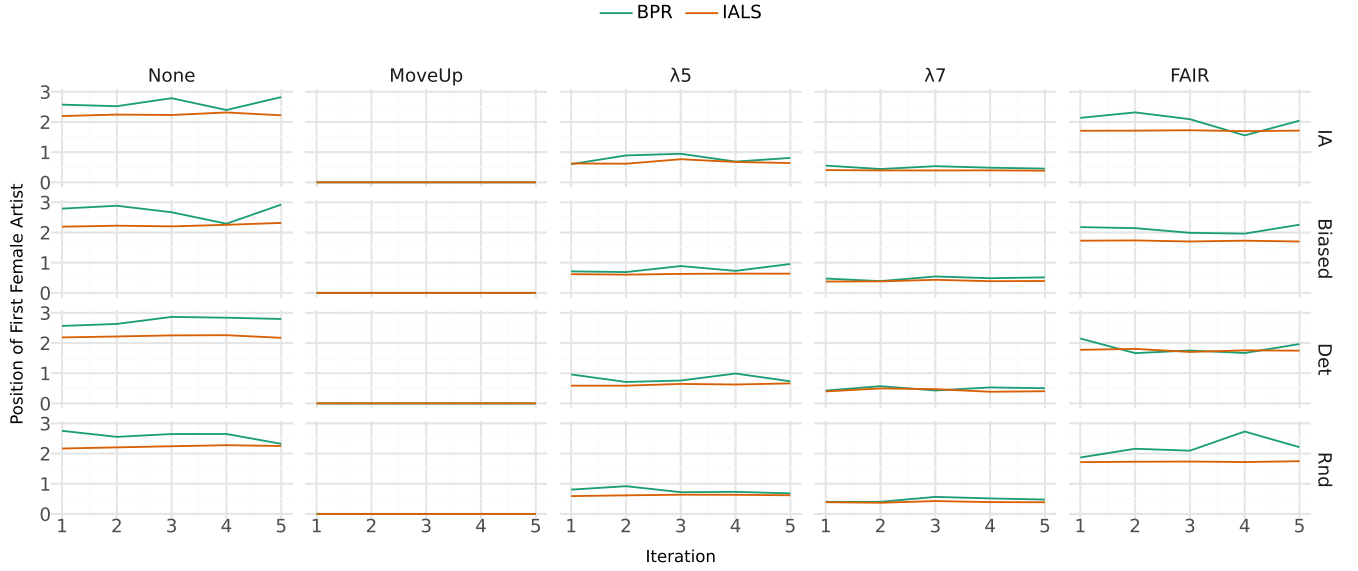


Figure 1: Average position of the first female artist (*PFA*) over iterations. Position 0 is the highest rank.

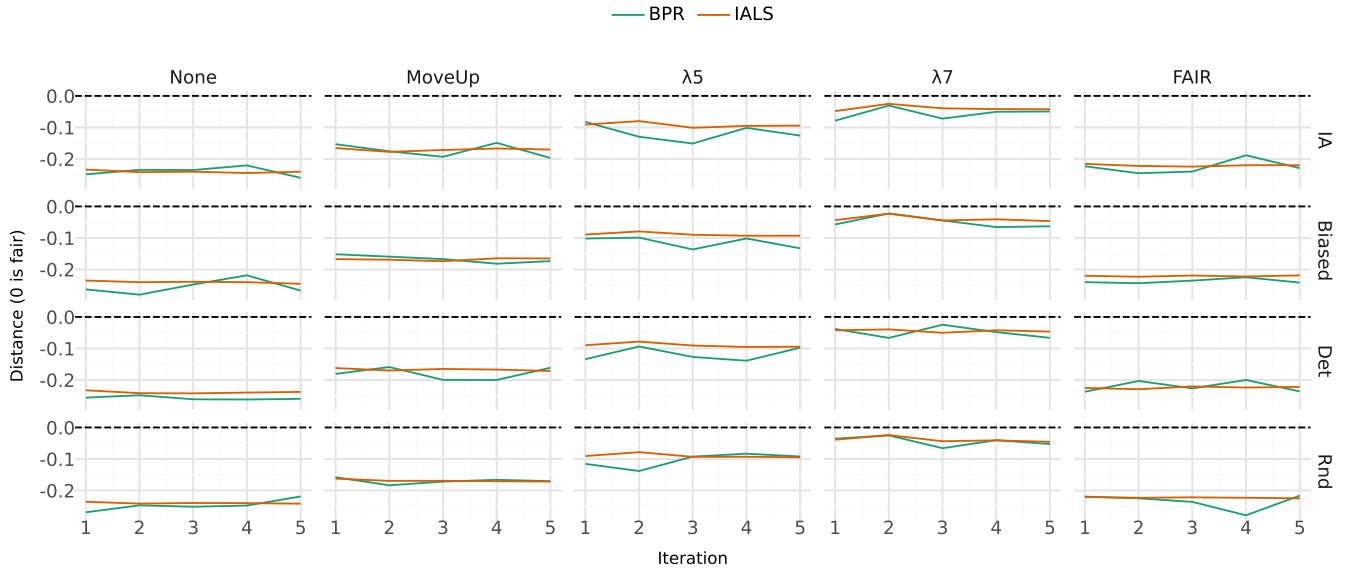


Figure 2: *AWRF* over iterations. 0 is fair; negative values mean that unprotected-group (male) artists are overexposed.

- *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

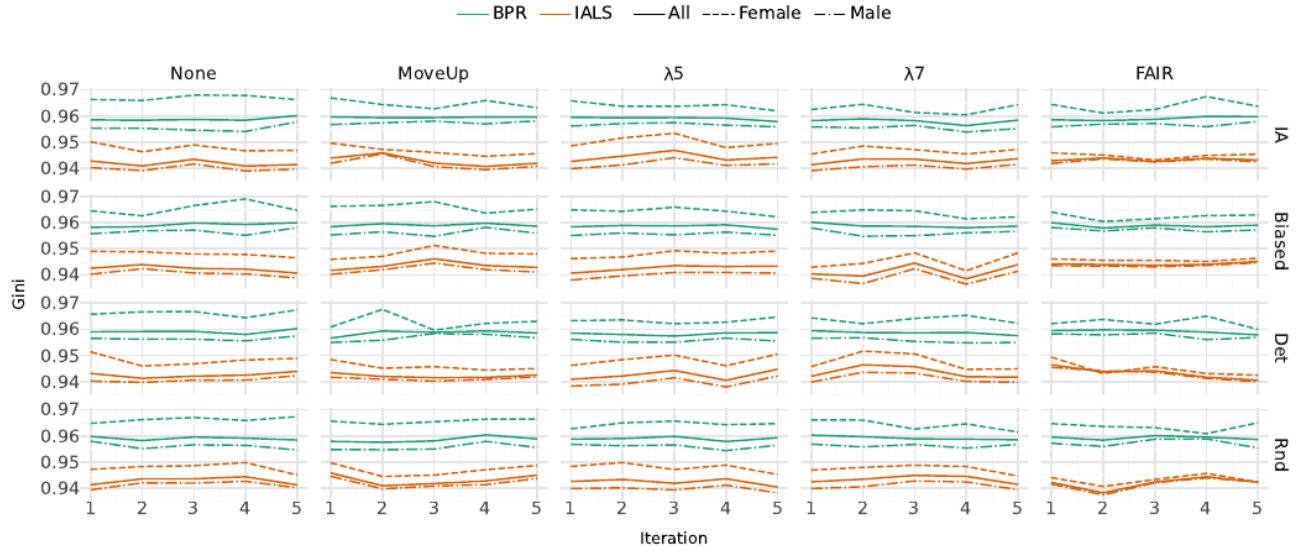


図3: 全体と男女アーティストの反復におけるジニ係数(Gini@10)。1 は最大不等式である。

Gini@k = 1は、すべての推薦が同じアーティストになることを示し、0は、すべてのアーティストが等しく推薦されることを意味する。

3.4 ユーザー選択モデル

シミュレーションでは、4つの異なるユーザー選択モデル(それぞれN = 10で実装)を検討する:

決定論的(Det) ユーザーはN個の推奨アイテムをすべて聞く。ランダム(Rnd) ユーザーの上位N個の推奨アイテムに含まれる各アイテムについて、ユーザーは確率0.5でリスニングを行う。InspectionAbandon (IA) 確率的カスケードモデル[27]: トップN推薦の各アイテムについて、確率0.5で、ユーザーはそのアイテムを聴く。リスニングまたは無視した後、彼らは確率0.3で完全にリスニングを停止し、そうでなければ次のアイテムにリスニングを続ける。IAモデルは、既存の評価文献[27]のカスケードモデルに基づいており、ユーザーは下位のものよりも上位の項目を消費する確率が高いことを実装している。

IAのバイアスされたバリエーションで、ユーザーが男性アーティストを聴く可能性が他の人より10%高くなるようにバイアスされている。

3.5 シミュレーション

これらの緩和策とユーザー選択モデルの長期的な効果を研究するために、ユーザー・推薦者フィードバックループをシミュレートする。各ユーザーについて、選択モデルを使用して、システムの推奨事項からアーティストを選択し、選択されたアーティストの追加リスニングイベントを記録する。次に、この拡張されたデータでモデルを再トレーニングし、次の反復のための推奨を計算する。先行研究[16]では5回繰り返した後に効果が固定されているため、この手順を5回繰り返す。

4 RESULTS

本節では、女性初心者(PFA)の位置、AW RF、ジニ係数に関する結果を示す。再順位付けアプローチ(図中のファセットプロットの列)とユーザー選択モデル(ファセットプロットの行)を比較する。データ、モデル、シミュレーションに完全なジェンダースペクトルを含めることは重要であるが、我々のデータには、分析に含めることから科学的結論を導き出すのに十分な非二元的アーティストが含まれていないため、男性アーティストと女性アーティストの結果のみを報告する。

図1は、再順位付け戦略が女性初心者の位置を決定するのに対し、ユーザー選択モデルは女性初心者の順位付けに影響を与えないことを明確に示している。IALSはBPRよりも安定しており、反復や選択モデルによるばらつきはほとんどないことがわかる。MoveUpは常に女性アーティストを一番最初のポジションに置くので、デザインによってこのメトリックを最適化する。λ7-次いでλ5-は、女性アーティストとして初めてわずかに低い順位でランク付けされる。FAIRは、最初の女性アーティストを、再ランク付けなしでベースラインよりわずかに高いだけランク付けする(なし)。

AW RFはPFAとほぼ一致する。図2より、再順位付け戦略はユーザー選択モデルよりもAW RFに大きな影響を与え(列間のばらつきが行よりも多い)、時間の経過による変化はほとんどないことがわかる。さらに、FAIRはベースラインと比較して、AW RFの改善に最も寄与しないことが観察された。最初の女性アーティスト(MoveUp)のみを動かした場合、単一アイテムの位置を調整するだけであるにもかかわらず、全体的な露出の公平性においてより効果的である。λ-re-rankerが最も効果的で、λ7が最も露出の公平性を向上させた。PFAと同様に、IALSに基づく推奨は、より安定したジェンダー暴露バランスを有していた。

図3は、Gini@10で測定した多様性を示している。IALSは、すべての再順位付け戦略と選択モデルにおいて、BPRよりも公平である。さらに、Gini@10_{female}とGini@10_{male}を比較すると、

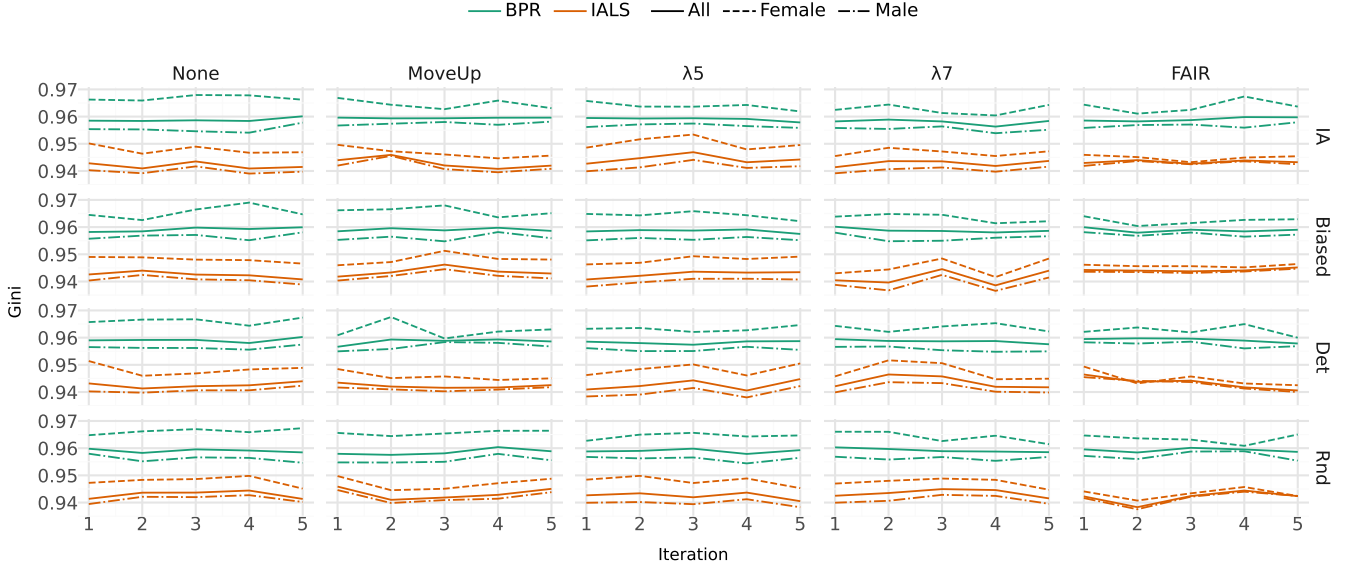


Figure 3: Gini coefficients ($Gini@10$) over iterations, both overall and for male and female artists. 1 is maximum inequality.

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

女性アーティストよりも男性アーティストの方が公平な結果を示している。女性アーティストが推奨されていないだけでなく(AW RF参照)、女性アーティストに接する機会は、それらのアーティストのうち、より少ない割合に集中している。FAIRでは、特にIALSベースモデルで不公平感の男女格差が縮まる。

5 考察と結論

我々の結果は、モデルが時間の経過とともに新しいユーザーとのインタラクションで再トレーニングされるにつれて、推薦のプロバイダーの公平性に対して、再ランク付け戦略がユーザー選択モデルよりも大きな効果を持つことを示している: 縦断的な公平性に関しては、「それはあなたではない」(ユーザー)、「それは私だ」(推薦システム)。この効果は、複数の指標と基礎となる推薦モデルで一貫している。つまり、ユーザーの選択と推薦への反応のモデルは、我々が検討した公平性の指標にほとんど影響を与えなかった。したがって、再ランク戦略は、アルゴリズムがアイテムの選択においてユーザーの行動よりも強い影響を与えるため、偏った世界に介入し、「社会的な不均衡」[26]に対処するための有用なツールとなり得る。我々が検討した基本推薦モデルは異なる挙動を示し、IALSはより安定した結果を提供し、BPRは明確な傾向はないものの、公平性と多様性の指標に大きなばらつきを示した。再順位付け戦略は元のフィードバックループを破ることができるが、IALSは我々の研究で観察された5回の反復の中で「新しい」フィードバックループを固定しているようである。

BPRのボラティリティがどの程度ユーザーデータ対ノイズへの応答であり、トレーニングにおけるランダム性への感度であるかは明らかではない。これらの効果を分離するためには、シミュレーションの再実行を繰り返す必要があり、計算コストがかかる。ノイズ源を特定するサンプル効率の良い方法[7]を開発することは、推薦システムシミュレーションの重要な将来の研究方向である。この研究を拡張し、改善するためのいくつかの重要な方向性がある。一つは、ノンバイナリーのジェンダー・アイデンティティを考慮したが、ノンバイナリーのアーティストの数が少ないため、現在の手法で意味のある分析をすることは難しい。今後、再ランキング戦略や選択モデル(より偏ったモデルを含む)、データセットやドメインの追加、推薦戦略(コンテンツベースやハイブリッドモデルを含む)など、より広範な戦略を模索する予定である。また、今後、疎外された複数のグループに属するアーティストに対する交差的な影響を検証することも重要である。

謝辞(謝辞)

本書は、ランド・ザルツブルグの助成金(助成番号20204-WISS/263/6-6022)によるデジタル科学と学際的技術の卓越性(EXDIGIT)プロジェクト、および米国の支援を受けた。全米科学財団(National Science Foundation)の助成金番号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] Oscar 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/10.1561/15000000079>
- [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, Soroush 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 Choueiri, 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>

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] Oscar 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 Dinissen 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/10.1561/15000000079>
- [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, Soroush Saghari, Shiva Parsa Rad, and Marko Tkalčić. 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) (wsHCMIIR '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 Abdollahpour, 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 Attention on 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 Eskandarian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-aspect Fairness through Personalized 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 Eskandarian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-aspect Fairness through Personalized 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] Jingjing 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>

- [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 Lacić. 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 Abdollahpour, 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 Attention on 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 Eskandarian, 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 Eskandarian, 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 Zehlke, 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] Jingjing 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>