Towards a Stakeholder-Centered Auditing Framework for Assessing Fairness in (Music) Recommender Systems

Valdy Wiratama
Department of Artificial Intelligence and Human Interfaces
University of Salzburg
Salzburg, Austria
valdy.wiratama@plus.ac.at

Abstract

Recommender systems (RS) play a central role in everyday activities by generating personalized recommendations that support users' choices. In this regard, evaluating whether these recommendations are fair is crucial. However, this is challenging because fairness criteria vary across stakeholders and existing approaches span many, often incompatible conceptions and operationalizations of metrics, as well as trade-offs. To address this problem, we propose a stakeholder-centered auditing framework that elicits and formalizes stakeholder perspectives on fairness in RS. We focus on the music domain, where artists and other item providers, listeners, and music streaming services (MSS) have potentially conflicting goals. Furthermore, we follow a three-stage mixed-methods approach: (1) data collection through literature review and semi-structured interviews; (2) framework development through the analysis and synthesis of collected data; and (3) evaluating the framework through scenario-based evaluations with stakeholders. Through this approach, we expect the outcome of a structured auditing approach that improves fairness assessments by redefining what is considered to be appropriate measures to capture fairness and its trade-offs.

CCS Concepts

• **Information systems → Recommender systems**; *Evaluation of retrieval results.*

Keywords

Algorithmic auditing, fairness, mixed methods, music streaming

ACM Reference Format:
Valdy Wiratama. 2026. Towards a Stakeholder-Centered Auditing Framework for Assessing Fairness in (Music) Recommender Systems. In *34th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '26), June 08–11, 2026, Gothenburg, Sweden.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3774935.3803067

1 Introduction

Recommender systems (RS) play a central role in enabling interactive processes through which users can influence system behavior and consequently facilitate algorithmic decision-making by generating personalized recommendations that support user choice in consumption, media engagement, and information access [8]. However, they can also raise significant challenges such as fairness. Fairness itself has emerged as a critical yet unresolved problematic behavior [9] and it concerns the absence of prejudice or favoritism toward individuals or groups [31].

When applied to RS, fairness is frequently evaluated using predefined technical proxies and static metrics that rely on implicit assumptions about acceptable distributions and outcomes [6]. Such an approach risks producing mathematically optimized objectives that do not correspond to the perception of fairness in real-world contexts, while creating an "abstraction trap," where formal metrics fail to reflect stakeholder experiences and values [9]. In multistakeholder environments, the failure of such becomes a limitation since RS affect diverse groups and cannot be evaluated solely based on the perspective of a single stakeholder [7].

Acknowledging the limitation, recent research has begun adopting auditing frameworks. Examples include a causal framework that uses peer comparisons to approximate counterfactual fairness [15] and a multi-objective framework that examines trade-offs between accuracy, user fairness, and provider fairness in RS [6]. However, these frameworks do not provide an integrated and operational approach for translating multiple stakeholder perspectives and interests on fairness into auditable criteria, measurements, and interpretive procedures for RS. Given this research gap, we propose the following key research question: **How can an auditing framework for recommender systems be established using a multifaceted articulation of stakeholder perspectives and interests on fairness?**

To address this question, we target the problem in the music domain for two reasons. First, given the subject of this research, narrowing down the scope is a necessity for producing context-sensitive and meaningful results. Fairness is not a one-size-fits-all value [9] and it holds divergent interpretations across various contexts [21]. Furthermore, RS are embedded in domain-specific ecosystems in which relevant fairness criteria, affected stakeholders, and socio-technical interactions vary [32]. Second, music is chosen because the domain is a multistakeholder environment that is characterized by a particularly biased starting point with respect to, for instance, gender and fair exposure or visibility of artists [17].

While music is used to narrow the scope, our research is expected to provide three contributions beyond the domain. First, the development of a stakeholder-centered auditing framework makes the integration of affected stakeholders' values and goals explicit. Second, our work provides a structured auditing approach that improves

UMAP '26, Gothenburg, Sweden

ACM ISBN 979-8-4007-2311-7/26/06
https://doi.org/10.1145/3774935.3803067

fairness assessments by refining approaches to fairness, addressing fairness trade-offs, and incorporating stakeholder-centered evaluations. Third, our work advances the field of personalization by accounting for normative considerations in fairness.

2 Related Work

2.1 Fairness in the Music Domain

Within the music domain, the digital music value chain involves multiple stakeholders with varying goals and interests [3]. Among these stakeholders are item providers (e.g., artists and record labels), listeners or end users, and music streaming services (MSS) [10]. MSS like Spotify and Apple Music help end users explore extensive catalogs of songs, which had reached a total of 253 million globally by the end of 2025 [24]. To help end users navigate this abundance and match them with songs, artists, and/or playlists that can cater to individual preferences, most MSS employ music recommender systems (MRS).

While MRS create more personalized listening experiences [29], the sole optimization on specific metrics (e.g., user satisfaction [10]) can lead to fairness challenges. For end users, these challenges include discrimination against a given group defined by a protected feature (e.g., gender or age) [4] and cases in which recommendation quality is significantly better for those with mainstream taste [22]. Likewise, challenges in provider fairness may occur when MRS, for instance, privilege the items of a small group of artists when maximizing user satisfaction [18], or when feedback-loop dynamics over repeated cycles contribute to decreased exposure of items, with consequences such as less revenue for item providers [16].

2.2 Fairness in Recommender Systems

A widely acknowledged way to define fairness in RS is from a distributional perspective. From this view, fairness concerns whether the benefits and resources provided by a system are allocated among affected parties, as well as how a system's positive and negative effects are distributed across its subjects [12]. Because of this structurally nuanced definition, assessing fairness is traditionally framed around "fairness for whom" [9]. Simultaneously, this framing distinguishes between consumer fairness and provider fairness. Consumer fairness concerns how systems affect end users and whether these effects are fair or lead to unjust harms [13], while provider fairness evaluates whether exposure is allocated fairly among item creators, who benefit when users discover their items [12].

Furthermore, current research in RS frequently focuses on fairness metrics, such as disparate exposure, ranking-based statistical parity, group utility disparities, and prediction error parity, in offline evaluations [9]. It is also commonly divided into studies that look into group fairness and individual fairness. The former is frequently instantiated through exposure- and ranking-based constraints [34], while the latter is expressed in ranking settings through formulations that evaluate each candidate separately [37] and with representational commitments that require similar individuals to be treated similarly [11].

Despite this apparent clarity in metrics and division, the landscape of fairness assessments contains dozens of metrics that are often incompatible [2]. Technical work also embeds fairness into optimization by adding regularizers and formalizing relevance–fairness trade-offs as multi-objective control problems [38]. Subsequently, this formalization is articulated in studies that focus on exposure-based ranking as an objective for balancing the items being ranked with the utility that the rankings provide to end users [34].

In domains like music, the objective is no different. Most discussions focus on evaluating fairness through exposure inequities. For example, Kowald et al. [23] show that MRS may produce recommendation lists that are extremely concentrated on popular items, even if a user prefers long-tail, non-popular items. Additionally, music items are embedded in discussions of cultural interpretation, which further complicate fairness assessments, since recommendation decisions can be subjective and carry ideological or social connotations that challenge purely metric-driven evaluations [16].

2.3 Auditing Recommender Systems

Existing work on auditing RS draws on broader discussions of algorithm audits, whose foundations lie in earlier audit studies from the field of social sciences. Within this field, algorithm audits are introduced as a way to examine opaque algorithms on internet platforms [30]. They are also formalized as a method of repeatedly and systematically querying an algorithm with inputs, while observing the corresponding outputs in order to draw inferences about its opaque inner workings [27].

Beyond the procedural definitions, algorithm audits are also structured around questions of affected actors and governance [19]. Users are considered as one of these actors because they can detect, understand, and interrogate problematic behavior through their interactions with algorithmic systems [33]. In addition, this consideration aligns with the noninvasive user audit tradition, which informs the foundation of our research, particularly because the tradition allows for the collection of information about users' normal interactions with a platform in order to infer how its algorithm operates [30].

Furthermore, the "Algorithmic Auditing for Music Discovery" (AA4MD)[1] research project [28] provides the closest point of reference for our research project. AA4MD looks at the relationship between RS' mechanisms and music discoverability, particularly in relation to the promotion of cultural diversity. However, instead of observing discoverability, our research project focuses on fairness, while simultaneously maintaining the emphasis on prioritizing the perspectives of those who are most affected by RS.

2.4 Stakeholder Inclusion in Auditing for Fairness in Recommender Systems

In RS, users are simultaneously producers of algorithmic input and consumers of algorithmic output, which makes them uniquely positioned to detect and reflect on algorithmic harm [19]. However, referring to these actors as "users" may be misleading, since fairness concerns in multistakeholder environments extend beyond end users. Using the term "stakeholders" is preferable, as it acknowledges that different actors who are affected by RS may hold distinct values and goals that correspond to different fairness needs [7]. Subsequently, this acknowledgment requires the inclusion of both end users and item providers, attention to how benefits (e.g., exposure or effectiveness) are distributed across stakeholders, and

[1] Project website: https://aa4md-project.eu/

a departure from the assumption that a single quantitative metric suffices in RS assessments [2].

In a related vein, quantitative metrics in fairness assessments frequently leave out normative claims and in-depth contextual meanings [9]. User studies also show that behavioral shifts toward quantitatively assessed fairness objectives may not translate into appropriate fairness judgments unless explanatory information is provided [1]. Even in ranking-based evaluations, fairness is also experience-sensitive because exposure is largely determined by presentation choices that mediate stakeholder outcomes and perceptions through interactions [34].

Considering the experience-sensitive nature and its dependence on interaction dynamics, an auditing framework for assessing fairness must be attentive to the conditions through which recommender effects are experienced among affected stakeholders. Meaningful audits and the development of an auditing framework also require attention to stakeholder interests, as their expressed priorities may help identify harms that experts might miss [26].

3 Methods

To develop a stakeholder-centered auditing framework for assessing fairness in RS, our doctoral research project adopts a mixed-methods approach. This approach is organized in three sequential stages. Stage 1 involves data collection to establish conceptual and empirical foundations. Meanwhile, Stage 2 proceeds with framework development by analyzing and synthesizing the collected data using qualitative and quantitative methods. Finally, Stage 3 involves applying and testing the developed framework by using stakeholder-centered scenario-based evaluations to ensure that the framework remains aligned with stakeholder interpretations.

Considering the focus on data collection in Stage 1, we plan to begin by conducting a literature review. Conducting this review allows for a synthesis of concepts, assumptions, and qualitative comparisons, which can also be used for the subsequent interview sessions. For these sessions, they are planned to be recorded and their structure integrates two components: a semi-structured format and a card-sorting activity. The latter is included to organize, categorize, and reflect on relevant expressions, as well as to identify what participants consider "most important" and "least important" [25]. After the sessions are completed, the recorded interviews are planned to be transcribed before moving on to Stage 2.

In Stage 2, the focus is on data analysis, where appropriate methods are applied to interview data (e.g., thematic analysis [5]) and card-sorting outcomes (e.g., clustering analysis [14] and dominance estimation [20]). Once the analysis is finished, the results are synthesized to produce an initial framework as a starting point.

Moving on to Stage 3, the initial framework from Stage 2 is tested through stakeholder-centered scenario-based evaluations. These evaluations are intended to allow further refinement of the framework through additional participant input (e.g., via interpretive formative evaluations [36] with stakeholders) and to achieve ecological validity in interpreting fairness [9].

4 Progress and Results

Figure 1 illustrates the planned and ongoing research activities using a Gantt chart. Within this chart, we grouped all the activities

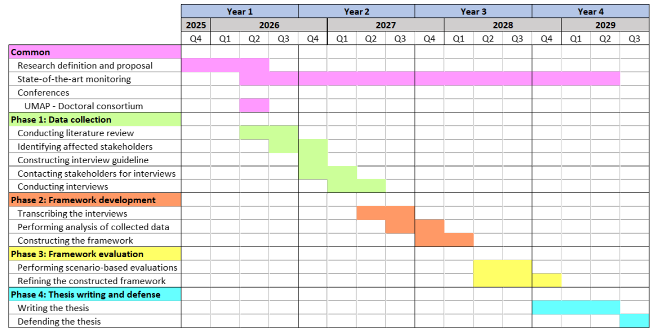

Figure 1: Doctoral research timeline as of Q1 2026

into five categories. The first category outlines the common activities that may occur throughout a doctoral research program. These include research definition and proposal development during the first three quarters, as well as the ongoing monitoring of the state of the art in RS, platform policies, and related industry changes.

At the moment, the research definition activity is complete and a proposal that outlines the research problem, methodology plan, proposed timeline, and alternatives is being finalized. As for the next activities, they correspond to Phase 1, which begins with conducting a literature review and is currently being prepared. Meanwhile, Phase 2 involves activities related to framework development. Phase 3 concerns the activities for evaluating the developed framework, while Phase 4 focuses on synthesizing the research, as well as writing it up and defending it as a doctoral thesis.

Looking beyond the planned phases, this research project is intended to result in a framework that makes auditing for fairness more actionable in practice. In this sense, the broader value of our project lies in its potential to support more structured and context-sensitive fairness assessments in RS.

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.

References

[1] Gabrielle Alves, Dietmar Jannach, Rodrigo Ferrari De Souza, and Marcelo Garcia Manzato. 2024. User Perception of Fairness-Calibrated Recommendations. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization* (Cagliari, Italy) *(UMAP '24)*. ACM, New York, NY, USA, 78–88. doi:10.1145/3627043.3659558

[2] Enrique Amigó, Yashar Deldjoo, Stefano Mizzaro, and Alejandro Bellogín. 2023. A Unifying and General Account of Fairness Measurement in Recommender Systems. *Information Processing & Management* 60, 1 (2023), 103115. doi:10.1016/j.ipm.2022.103115

[3] Christine Bauer and Eva Zangerle. 2019. Leveraging Multi-Method Evaluation for Multi-Stakeholder Settings. In *Proceedings of the 1st Workshop on the Impact of Recommender Systems co-located with the 13th ACM Conference on Recommender Systems (ImpactRS@RecSys 2019)* (Copenhagen, Denmark) *(CEUR Workshop Proceedings, Vol. 2462)*. CEUR-WS.org. Retrieved February 25, 2026 from https://ceur-ws.org/Vol-2462/short3.pdf

[4] Ludovico Boratto, Gianni Fenu, Mirko Marras, and Giacomo Medda. 2022. Consumer Fairness in Recommender Systems: Contextualizing Definitions and Mitigations. In *Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10-14, 2022, Proceedings, Part I (Lecture Notes in Computer Science)*. Springer, 552–566. doi:10.1007/978-3-030-99736-6_37

[5] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/

1478088706qp063oa

[6] Salvatore Bufi, Vincenzo Paparella, Vito Walter Anelli, and Tommaso Di Noia. 2025. Legal but Unfair: Auditing the Impact of Data Minimization on Fairness and Accuracy Trade-off in Recommender Systems. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2025, New York City, NY, USA, June 16-19, 2025* (New York, NY, USA). ACM, 114–123. doi:10.1145/3699682.3728356

[7] Robin Burke, Gediminas Adomavicius, Toine Bogers, Tommaso Di Noia, Dominik Kowald, Julia Neidhardt, Özlem Özgöbek, Maria Soledad Pera, Nava Tintarev, and Jürgen Ziegler. 2025. De-centering the (Traditional) user: Multistakeholder evaluation of recommender systems. *International Journal of Human-Computer Studies* 203 (2025), 103560. doi:10.1016/J.IJHCS.2025.103560

[8] Robin D. Burke. 2007. Hybrid Web Recommender Systems. In *The Adaptive Web, Methods and Strategies of Web Personalization (Lecture Notes in Computer Science)*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). Springer, 377–408. doi:10.1007/978-3-540-72079-9_12

[9] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogín, Alessandro Difonzo, and Dario Zanzonelli. 2024. Fairness in Recommender Systems: Research Landscape and Future Directions. *User Modeling and User-Adapted Interaction* 34, 1 (2024), 59–108. doi:10.1007/S11257-023-09364-Z

[10] Karlijn Dinnissen and Christine Bauer. 2022. Fairness in Music Recommender Systems: A Stakeholder-Centered Mini Review. *Frontiers Big Data* 5, Article 913608 (2022), 9 pages. doi:10.3389/FDATA.2022.913608

[11] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. 2012. Fairness through awareness. In *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, Shafi Goldwasser (Ed.). ACM, 214–226. doi:10.1145/2090236.2090255

[12] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Human–Computer Interaction* 16, 1-2 (2022), 1–177. doi:10.1561/1500000079

[13] Michael D. Ekstrand and Maria Soledad Pera. 2022. Matching Consumer Fairness Objectives & Strategies for RecSys. *CoRR* abs/2209.02662 (2022). arXiv:2209.02662 doi:10.48550/arXiv.2209.02662

[14] Angi Faiks and Nancy Hyland. 2000. Gaining User Insight: A Case Study Illustrating the Card Sort Technique. *College & Research Libraries* 61, 4 (2000), 349–357. doi:10.5860/crl.61.4.349

[15] Shiqi Fang, Zexun Chen, and Jake Ansell. 2024. Peer-induced Fairness: A Causal Approach for Algorithmic Fairness Auditing. *CoRR* abs/2408.02558 (2024). arXiv:2408.02558 doi:10.48550/arXiv.2408.02558

[16] 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 (wsHCMIR '19)* (Delft, The Netherlands). 8–10. doi:10.48550/arXiv.1911.04827

[17] 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 *Proceedings of the 18th ACM Conference on Recommender Systems, RecSys 2024, Bari, Italy, October 14-18, 2024* (Bari, Italy). ACM, 884–889. doi:10.1145/3640457.3688163

[18] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the Loop: Gender Imbalance in Music Recommenders. In *CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021*. ACM, 249–254. doi:10.1145/3406522.3446033

[19] Uri Y. Hacohen. 2025. User-based Algorithmic Auditing. *Theoretical Inquiries in Law* 26, 1 (2025), 287–311. doi:10.1515/til-2025-0012

[20] Daniel A. Henderson. 2025. Modelling and Analysis of Rank Ordered Data with Ties via a Generalized Plackett-Luce Model. *Bayesian Analysis* 20, 3 (2025), 1109–1137. doi:10.1214/24-BA1434

[21] Antoine Henry, Valdy Wiratama, Adelaida Afilipoaie, Heritiana Ranaivoson, and Eric Arrivé. 2024. Impacts of AI on Music Consumption and Fairness. *Emerging Media* 2, 3 (2024), 382–396. doi:10.1177/27523543241269047

[22] Dominik Kowald, Peter Müllner, Eva Zangerle, Christine Bauer, Markus Schedl, and Elisabeth Lex. 2021. Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science* 10, 1 (2021), 14. doi:10.1140/EPJDS/S13688-021-00268-9

[23] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study. In *Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II (Lecture Notes in Computer Science)*, Joemon M. Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J. Silva, and Flávio Martins (Eds.). Springer, 35–42. doi:10.1007/978-3-030-45442-5_5

[24] Luminate. 2026. 2025 Year-End Music Report. https://luminatedata.com/reports/yearend-music-industry-report-2025 Access requires submitting user details. Accessed February 24, 2026.

[25] Michal Luria. 2023. Co-Design Perspectives on Algorithm Transparency Reporting: Guidelines and Prototypes. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2023, Chicago, IL, USA, June 12-15, 2023*. ACM, 1076–1087. doi:10.1145/3593013.3594064

[26] Anna-Katharina Meßmer and Martin Degeling. 2023. Auditing Recommender Systems – Putting the DSA into Practice with a Risk-scenario-based Approach. *CoRR* abs/2302.04556 (2023). arXiv:2302.04556 doi:10.48550/arXiv.2302.04556

[27] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff T. Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Found. Trends Hum. Comput. Interact.* 14, 4 (2021), 272–344. doi:10.1561/1100000083

[28] Lorenzo Porcaro, Emilia Gómez, and Tiziana Catarci. 2024. End-user Algorithmic Auditing for Music Discoverability: A Research Roadmap. In *Proceedings of the 2nd Music Recommender Systems Workshop 2024 co-located with the 18th ACM Conference on Recommender Systems (RecSys 2024)* (Bari, Italy) *(CEUR Workshop Proceedings, Vol. 3787)*. CEUR-WS.org. Retrieved February 18, 2026 from https://ceur-ws.org/Vol-3787/paper4.pdf

[29] Jing Ren, Robert J. Kauffman, and Dave King. 2019. Two-sided Value-based Music Artist Recommendation in Streaming Music Services. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (Maui, HI, USA) *(Proceedings of the Annual Hawaii International Conference on System Sciences)*. Hawaii International Conference on System Sciences (HICSS), 2679–2688. doi:10.24251/HICSS.2019.323

[30] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*. Seattle, WA, USA, 1–23. https://www.kevinhamilton.org/share/papers/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

[31] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. 2019. How Do Fairness Definitions Fare?: Examining Public Attitudes Towards Algorithmic Definitions of Fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*. ACM, 99–106. doi:10.1145/3306618.3314248

[32] Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, GA, USA, January 29-31, 2019*. ACM, 59–68. doi:10.1145/3287560.3287598

[33] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 433:1–433:29. doi:10.1145/3479577

[34] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*. ACM, 2219–2228. doi:10.1145/3219819.3220088

[35] Nasim Sonboli, Jessie J. Smith, Florencia Cabral Berenfus, Robin Burke, and Casey Fiesler. 2021. Fairness and Transparency in Recommendation: The Users' Perspective. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2021, Utrecht, The Netherlands, June, 21-25, 2021*. ACM, 274–279. doi:10.1145/3450613.3456835

[36] Cheryl B. Stetler, Marcia W. Legro, Carolyn M. Wallace, Candice Bowman, Marylou Guihan, Hildi Hagedorn, Barbara Kimmel, Nancy D. Sharp, and Jeffrey L. Smith. 2006. The Role of Formative Evaluation in Implementation Research and the QUERI Experience. *Journal of General Internal Medicine* 21, 2 (Feb. 2006), 51–58. doi:10.1007/s11606-006-0267-9

[37] Tao Yang, Zhichao Xu, and Qingyao Ai. 2023. Vertical Allocation-based Fair Exposure Amortizing in Ranking. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, SIGIR-AP 2023, Beijing, China, November 26-28, 2023*. ACM, 234–244. doi:10.1145/3624918.3625313

[38] Sirui Yao and Bert Huang. 2017. Beyond Parity: Fairness Objectives for Collaborative Filtering. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 2921–2930. https://proceedings.neurips.cc/paper/2017/hash/e6384711491713d29bc63fc5eeb5ba4f-Abstract.html