End-user Algorithmic Auditing for Music Discoverability: A Research Roadmap

Lorenzo Porcaro[1,2], Emilia Gómez[2] and Tiziana Catarci[1]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Italy
[2]Joint Research Centre, European Commission, Spain[‡]

Abstract

The evolution of online platforms over the past decades has radically transformed the way people discover music, and thanks to social media and music streaming services nowadays listeners have access to an ever-increasing amount of tracks and artists. Within these platforms, one of the goals of recommender systems is to help users discover music without making them feel overwhelmed while exploring the huge catalogues available. However, these systems have come under scrutiny from the scientific community, policy-makers, and civil society due to their potential negative societal impact, notably with regard to issues of fairness, non-discrimination, inclusion and diversity. Algorithmic auditing has emerged as a tool to analyse the problematic behaviours exhibited by recommenders, and to offer remedies that can limit their negative impact. In this position paper, we advocate for the involvement of end-users in the auditing process, which can contribute to the recognition, analysis, and mitigation of problematic behaviours which may arise while discovering music. Highlighting how recommenders, by influencing the discoverability of music, may impact listeners' exposure to culturally diverse content, we seek to address the challenges posed by music recommender systems problematic behaviours, ultimately with the goal of fostering a more inclusive and diverse environment for music discovery within the digital landscape.

Keywords

Human-Computer Interaction, Music Information Retrieval

1. Introduction

The music sector is one among the Cultural and Creative Sectors (CCSs) with the largest audience reach, and an essential component of cultural diversity. It has the power to bring positive changes in society and people's well-being, whilst at the same time generating billions of revenue. Among the challenges this sector is facing [1], the digital shift witnessed in the last decades has produced a radical change in the way we discover music. Tellingly, today's music consumption is enormously affected by the widespread adoption of online platforms such as social media, video-sharing and streaming services.

In these platforms, recommender systems help listeners discover new music within the huge catalogues available. At the same time, these systems benefit the artists, who on online platforms seek a connection with the most appropriate audience. However, the growing enthusiasm accompanying the rise of online platforms and recommender systems has been followed by increasing concerns about emerging problematic behaviours and their impact on society. In particular, the negative impact on the exercise of fundamental rights, such as the right to non-discrimination and respect for cultural diversity, is of particular relevance for the CCSs [2]. Under this lens, we believe that it is essential to investigate the potential of algorithmic auditing as a means to assess the impact of online platforms' recommender systems on the discoverability of music content, as well as their role in promoting cultural diversity.

In what follows, we motivate our position by briefly examining background research on related topics. We then continue by discussing the research objectives and questions we consider as relevant, for then describing the methodology we designed to tackle the presented challenges. Finally, we conclude by highlighting the potential impact that such research activities could have in the future.

[‡]The views expressed are purely those of the author(s) and may not in any circumstances be regarded as stating an official position of the European Commission.

MuRS 2024: 2nd Music Recommender Systems Workshop, October 14th, 2024, Bari, Italy

lorenzo.porcaro@gmail.com (L. Porcaro); emilia.gomez@ec.europa.eu (E. Gómez); catarci@diag.uniroma1.it (T. Catarci)
0000-0003-0218-5187 (L. Porcaro); 0000-0003-4983-3989 (E. Gómez); 0000-0002-3578-1121 (T. Catarci)

2. Background

Our research roadmap builds upon the idea of *end-user* [3] (or *everyday* [4]) *algorithmic auditing* put forth by scholars who suggest that it can serve the following purposes: 1) identify instances of problematic behaviours resulting from normal algorithm usage, and raise awareness of such practices; 2) generate hypotheses regarding observed behaviour and conduct tests on algorithmic systems to gather additional evidence; and 3) facilitate change by addressing identified issues through investigations.

Another pillar of our research is the concept of *music discoverability*, which can be defined as "[...] its availability online and its ability to be found among a wide range of other content, particularly by someone who was not specifically looking for it" [5]. Algorithmic auditing has not yet been applied to recommender systems for investigating the link between music discoverability and the promotion of cultural diversity, and related research is still in its infancy [6, 7, 8]. Therefore, our understanding of this phenomenon and its possibilities and drawbacks remains significantly restricted. Under this lens, the development of a specific algorithmic auditing framework is necessary to inform future policies and regulations [9].

In the music sector, the recent report on the impact of algorithmically driven recommendations on music consumption and production commissioned by the UK Centre for Data Ethics and Innovation [10] underlines how findings are still fragmented between different disciplines, particularly with regard to the literature on bias, diversity, fairness and transparency. Additionally, the report on the impact of AI on cultural curation redacted by the Canadian Schwartz Reisman Institute [11] highlights that medium- and long-term consequences of cultural curation and distribution unleashed by algorithmic systems have until now largely been overlooked by researchers and policy-makers. Under this lens, the relationship between recommender systems' mechanisms, their role in mediating music discoverability, and their impact on listeners' interaction with culturally diverse content remains unclear.

Nonetheless, it has been shown how users perceive diversity as a motivating factor for discovering music in the long term [12], further proving that, with a careful design, recommender systems can serve as agents of positive attitudinal change. Taking this perspective into account, a concept arises: the notion of crafting recommender systems with the aim of fostering cultural citizenship [13], moving beyond the pursuit of personalization [14] that is frequently influenced by commercial motives.

3. Research objectives and questions

We identify three specific research objectives (ROs) and corresponding research questions (RQs) on which the investigation should focus on:

RO1. Algorithmic Auditing as a source of evidence: Initially, it is mandatory to examine instances of recommender systems' problematic behaviours in online platforms not limited to the music field, understanding how the lessons learned can be applied to the music sector. The related research question (RQ1) is: *What is the potential of auditing music recommender systems, and what obstacles must be overcome?* We hypothesise that through the identification of a feasible number of instances of problematic behaviours, and by documenting the lessons learned and their applicability to the music field, we can provide evidence of how recommender systems may be audited.

RO2. Towards a framework for auditing discoverability: Second, it is needed to develop a tool for end-user auditing, focusing on the online platforms where recommender systems mediate the discoverability of music content. The related research question (RQ2) is: *How can we audit recommender systems' mechanisms affecting music discoverability?* By evaluating the functionality and effectiveness of the developed tool in identifying and auditing recommender system mechanisms, we plan to verify, through testing and validation processes, its ability to accurately audit problematic behaviours.

RO3. Facilitating the discoverability of culturally diverse music: Lastly, it should be considered the use of audits as a tool for remediation, proposing solutions to facilitate the algorithmic-mediated discoverability of culturally diverse music, where needed. The associated research question (RQ3) is the following: *How can we design recommender systems for promoting cultural diversity, by limiting their*

problematic behaviours? We intend to verify the effectiveness of the proposed solutions by comparing the empirical evidence gathered with the state of the art in algorithmic auditing [15] and recommender systems evaluation [16, 17], particularly focusing on the music field and the promotion of culturally diverse content in online platforms.

4. Methodology

The methodology we propose is aligned with the aforementioned research objectives and research questions. We designed a two-phase mixed-method exploratory study using the outcomes of the initial qualitative phase to facilitate the development of the subsequent quantitative phase. Figure 1 depicts an overview of the proposed methodology, and the methods required to reach the aforementioned ROs are described hereafter.

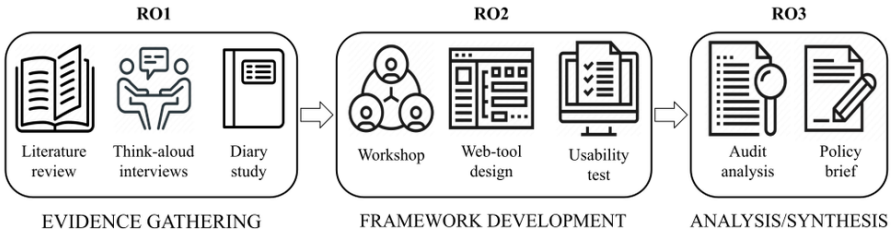

Figure 1: Overview of the proposed methodology with corresponding ROs. Additional information about each step of the methodology can be found in [18, 19, 20].

RO1 will involve the gathering of evidence on how algorithmic auditing can be applied to music recommender systems. We will start by reviewing existing cases of algorithmic auditing of recommender systems where problematic behaviours have been reported. This review will provide real-world scenarios that will be afterwards discussed with end-users, by means of think-aloud interviews. During the interviews, we will discuss existing cases of problematic behaviours of recommender systems, concerning topics such as bias, diversity, fairness and non-discrimination. Then, we will look together with the participants for new cases, performing open-ended tasks and asking them to think aloud to provide insights. Finally, we will invite them to look for cases of problematic behaviours in their everyday life while discovering music on online platforms, by taking part in a diary study. They will be asked to record such cases and provide an explanation of why they thought it would be problematic.

In this qualitative phase, recruiting participants for think-aloud interviews and the diary study might be challenging. Moreover, ensuring diverse and engaged participants is crucial for obtaining comprehensive insights. In order to tackle these issues, we will i) clearly communicate the purpose and benefits of participation, ii) target a diverse group of listeners, and iii) use incentives to motivate engagement.

RO2 will focus on the design and test of an algorithmic auditing tool for assessing the problematic behaviours which recommender systems may present whilst discovering music. The cases of problematic behaviour related to music discovery in online platforms submitted by participants during the diary study will be discussed in a workshop. Participants will work together to understand the identified cases by asking and answering questions about the reported issue. The analysis of the workshop's insights will serve as a basis for the next stage, wherein we will make use of crowd-based user-centred design techniques [21] for implementing a web-based tool for end-user auditing of recommender systems. This tool will empower users in conducting large-scale hypothesis testing on the problematic behaviours connected with music discoverability, and in generating audit reports to effectively communicate their findings. The tool will be evaluated in terms of i) usability, hence how effectively and satisfactorily it allows the users to achieve their goals, ii) user experience (UX), so the overall quality of the interaction with the tool, and iii) acceptance, for understanding the potential for future adoption of such tool.

In order to conduct a productive workshop where participants collaboratively discuss and understand identified cases of problematic behaviour, we will prepare well-structured materials, including clear case descriptions and discussion prompts. We will facilitate the workshop effectively, encouraging open communication, and moderating discussions to ensure productive collaboration. Moreover, evaluating the usability, user experience, and acceptance of the auditing tool requires careful planning to capture users' genuine interactions and perceptions.

RO3 will focus on the analysis and interpretation of the previously generated audit reports in order to inform policy-makers about the status of the discoverability of culturally diverse music in online platforms. Finally, we will review the audit reports generated throughout the research by the study participants using both qualitative methods (grounded theory) and quantitative analysis (text mining, statistical analysis). The end goal of the process will be to consolidate the accumulated knowledge in a policy brief.

Integrating qualitative methods and quantitative analysis can be complex and requires a coherent approach. We will develop a systematic process to combine insights, considering how qualitative findings can inform the formulation of quantitative research questions and vice versa. Furthermore, creating a comprehensive policy brief that effectively communicates the accumulated knowledge might be challenging.

A final remark needs to be made on the recruitment of participants. Indeed, based on ongoing empirical and theoretical analysis, it has been shown that women, LGBTQIA+ communities, and ethnic minorities often experience more impactful negative effects from problematic algorithmic behaviours [3, 15, 22]. In the whole research, we aim to conduct a thorough investigation into the underlying reasons for this trend. In fact, we argue that online platforms facilitating music discovery should proactively address this issue by ensuring better representation of marginalized communities. This will also mean reaching out to fellow researchers of such platforms to create a dialogue in order to make exploitable all the outcomes of the proposed research.

During the aforementioned empirical inquiries, we will actively seek to balance the sample group if we notice an overrepresentation of any other dominant group. While collecting data, we will only gather information relevant to the research, even if in certain cases we may include questions to ensure diversity and equal representation in the sample. This will prevent from relying solely on data from a homogenous group, ensuring a more comprehensive understanding of the issues. Nevertheless, working with a diverse group of participants may present challenges such as communication barriers, varying levels of technological literacy, and cultural differences. To address these, it will be pivotal to provide clear instructions, offer multilingual support when needed, and ensure accessibility for all participants. Additionally, fostering cultural sensitivity and establishing guidelines for conflict resolution can help create a respectful and inclusive environment, enabling effective collaboration among participants.

5. Scientific, societal and economic impacts

The multifaceted impact of the planned research will extend beyond its immediate duration, fostering positive changes in scientific discourse, policy-making, and social awareness, all while catalysing advancements in the understanding and utilisation of algorithmic auditing. The impact, extending beyond its immediate scope and duration, can be categorised into the following dimensions:

Scientific impacts. The approach to algorithmic auditing developed in this research activities will significantly contribute to the ongoing discourse concerning the inclusion of end-users in the identification of problematic algorithmic behaviours and the proposition of potential solutions. While human-centred design has long been a cornerstone in technology development, its application to address ambiguous mechanisms behind these technologies has received limited attention. We will fill this gap, sparking a new wave of exploration in this field. Scholars working in the fields of Human-Computer Interaction and Recommender Systems will mostly benefit from the research outcomes.

Economic and political impacts. The described strategies to advance the state of the art of recommender systems' auditing will hold particular significance within recent legislative frameworks,

e.g., the EU initiative on cultural diversity and the conditions for authors in the European music streaming market [23], the US Living Wage for Musicians Act [24], the UK inquiry on the economics of music streaming [25], and the Uruguay regulation of remuneration for literary works, or for the performance of musical works [26]. By enhancing the transparency of recommender systems employed in online platforms, the research outcomes will offer valuable insights to policy-makers, contributing to the creation of safer, transparent, and trustworthy online environments.

Social and cultural impacts. The promotion of a culturally diverse landscape in the music sector stands as a core objective for various initiatives advocated by several institutions. In this context, the contributions made by the presented research will be instrumental in realising this goal. The involvement of users in the critical analysis of the technology that underpins music discoverability aligns with this objective. By showing the inner workings of recommender systems, the research will empower users to enhance their algorithmic literacy further promoting informed digital engagement.

6. Final Remarks

The research roadmap described in this position paper is part of the European Commission's funded project **Algorithmic Auditing for Music Discovery (AA4MD)**, funded under the Marie Skłodowska-Curie Actions (MSCA) (Grant agreement ID: 101148443) [27].

References

[1] European Commission, Directorate-General for Communications Networks, Content and Technology, K. Izsak, A. Terrier, S. Kreutzer, T. Strähle, C. Roche, M. Moretto, S. Sorensen, M. Hartung, K. Knaving, M. Johansson, M. Ericsson, D. Tomchak, Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors, Publications Office of the European Union, 2022. doi:doi/10.2759/144212.

[2] European Parliament, Directorate-General for Internal Policies of the Union, B. Caramiaux, The use of artificial intelligence in the cultural and creative sectors – Concomitant expertise for INI report – Research for CULT Committee, European Parliament, 2020. doi:doi/10.2861/602011.

[3] M. S. Lam, M. L. Gordon, D. Metaxa, J. T. Hancock, J. A. Landay, M. S. Bernstein, End-user audits: A system empowering communities to lead large-scale investigations of harmful algorithmic behavior, Proc. ACM Hum.-Comput. Interact. 6 (2022). doi:10.1145/3555625.

[4] H. Shen, A. DeVos, M. Eslami, K. Holstein, Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors, Proc. ACM Hum.-Comput. Interact. 5 (2021). URL: https://doi.org/10.1145/3479577. doi:10.1145/3479577.

[5] Ministère de la Culture et des Communications du Québec, Ministère de la Culture de France., Mission franco-québéquoise sur la découvrabilité en ligne des contenus culturels francophones [Franco-Quebec mission on the online discoverability of French-language cultural content], Technical Report, Bibliothèque et Archives nationales du Québec, 2020. URL: https://www.vie-publique.fr/rapport/277472-rapport-sur-decouvrabilite-en-ligne-des-contenus-culturels-francophones.

[6] D. Lamprecht, M. Strohmaier, D. Helic, A method for evaluating discoverability and navigability of recommendation algorithms, Computational social networks 4 (2017) 1–26.

[7] L. R. Hanania, Streaming platforms and profiling – risks and opportunities for the discoverability of diversified cultural content, Medijske studije 13 (2022) 10–26.

[8] L. Briand, T. Bontempelli, W. Bendada, M. Morlon, F. Rigaud, B. Chapus, T. Bouabça, G. Salha-Galvan, Let's get it started: Fostering the discoverability of new releases on Deezer, in: European Conference on Information Retrieval, 2024, pp. 286–291.

[9] C. Magis, From "diversity" to "discoverability": Platform economy, algorithms and the transformations of cultural policies, in: Systemic Bias, Routledge, 2022.

[10] D. Hesmondhalgh, R. Campos Valverde, D. B. V. Kaye, Z. Li, The Impact of Algorithmically Driven Recommendation Systems on Music Consumption and Production: A Literature Review, Technical

Report, UK Centre for Data Ethics and Innovation Reports, 2023. URL: https://ssrn.com/abstract=4365916.

[11] G. Born, J. Morris, F. Diaz, A. Anderson, Artificial intelligence, music recommendation, and the curation of culture, Technical Report, The Schwartz Reisman Institute for Technology and Society (SRI), 2021. URL: https://tspace.library.utoronto.ca/handle/1807/129105.

[12] L. Porcaro, E. Gómez, C. Castillo, Assessing the impact of music recommendation diversity on listeners: A longitudinal study, ACM Trans. Recomm. Syst. 2 (2024). doi:10.1145/3608487.

[13] A. Ferraro, G. Ferreira, F. Diaz, G. Born, Measuring commonality in recommendation of cultural content to strengthen cultural citizenship, ACM Trans. Recomm. Syst. 2 (2024). doi:10.1145/3643138.

[14] B. P. Knijnenburg, S. Sivakumar, D. Wilkinson, Recommender systems for self-actualization, in: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 11–14. doi:10.1145/2959100.2959189.

[15] J. Bandy, Problematic machine behavior: A systematic literature review of algorithm audits, Proc. ACM Hum.-Comput. Interact. 5 (2021). doi:10.1145/3449148.

[16] E. Zangerle, C. Bauer, Evaluating recommender systems: Survey and framework, ACM Comput. Surv. 55 (2022). doi:10.1145/3556536.

[17] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer US, Boston, MA, 2015, pp. 309–352. doi:10.1007/978-1-4899-7637-6_9.

[18] U. Flick, The SAGE handbook of qualitative data analysis, SAGE Publications Ltd, 2013. doi:10.4135/9781446282243.

[19] J. Vanderdonckt, P. Palanque, M. Winckler, Handbook of Human Computer Interaction, Springer Cham, 2020. doi:10.1007/978-3-319-27648-9.

[20] M. Sienkiewicz, V. Sucha, Science for Policy Handbook, Elsevier, 2020. doi:10.1016/C2018-0-03963-8.

[21] T. Catarci, A. Marrella, G. Santucci, M. Sharf, A. Vitaletti, L. Di Lucchio, L. Imbesi, V. Malakuczi, From consensus to innovation. evolving towards crowd-based user-centered design, International Journal of Human–Computer Interaction 36 (2020) 1460–1475. doi:10.1080/10447318.2020.1753333.

[22] A. DeVrio, M. Eslami, K. Holstein, Building, shifting, & employing power: A taxonomy of responses from below to algorithmic harm, in: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 1093–1106. doi:10.1145/3630106.3658958.

[23] European Parlament Culture and Education Committee, Cultural diversity and the conditions for authors in the european music streaming market, 2023. URL: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2023/2054(INI).

[24] US Congress, H.R.7763 - Living Wage for Musicians Act of 2024, 2024. URL: https://www.congress.gov/bill/118th-congress/house-bill/7763/text.

[25] UK Department for Digital, Culture, Media & Sport (DCMS) Select Committee, Economics of music streaming, 2021. URL: https://committees.parliament.uk/work/646/economics-of-music-streaming/publications/.

[26] Uruguay Ministerio de Educación y Cultura, Decreto N° 404/023, de 12/12/2023. Reglamentación de Remuneraciones sobre Obras literarias, o de interpretación de Obras musicales., 2023. URL: https://medios.presidencia.gub.uy/legal/2023/decretos/12/mec_434.pdf.

[27] CORDIS, Algorithmic Auditing for Music Discoverability (AA4MD), 2024. doi:10.3030/101148443.