

# Surfacing Problematic Recommender System Behaviors Affecting Music Discoverability: A Think-Aloud Protocol

Lorenzo Porcaro  
Valeria Mirabella  
l.porcaro@diag.uniroma1.it  
mirabella@diag.uniroma1.it  
Department of Computer, Control,  
and Management Engineering  
Antonio Ruberti, Sapienza University  
of Rome  
Italy

Emilia Gómez\*  
emilia.gomez-  
gutierrez@ec.europa.eu  
Joint Research Centre,  
European Commission  
Music Technology Group, Universitat  
Pompeu Fabra  
Spain

Tiziana Catarci  
catarci@dis.uniroma1.it  
Institute of Cognitive Sciences and  
Technologies, National Research  
Council  
Italy  
Department of Computer, Control,  
and Management Engineering  
Antonio Ruberti, Sapienza University  
of Rome  
Italy

## Abstract

Recommender systems are central to contemporary music listening, yet their problematic behaviors remain underexplored from the perspective of everyday listeners. While prior research has addressed issues such as bias and diversity, less is known about how users themselves perceive and interpret these dynamics in relation to music discoverability. This paper reports on think-aloud interviews with 20 Italian digital-native listeners, who completed discovery-oriented tasks while reflecting on algorithmic recommendations. Thematic analysis revealed three recurring concerns: reinforcement of societal biases, commercial imperatives driving exposure, and confinement within narrow niches. These findings show how listeners actively develop folk theories of recommender behavior, highlighting a tension between algorithmic efficiency and cultural effects. We contribute empirical insights into user sensemaking of algorithmic harms, consolidate the use of the Think-Aloud Protocol as a user-driven auditing method, and outline design implications for more participatory and equitable music recommender systems.

## CCS Concepts

• **Human-centered computing** → Empirical studies in HCI; Empirical studies in collaborative and social computing; • **Information systems** → Recommender systems.

## Keywords

algorithmic auditing, music streaming platforms, music information retrieval

## ACM Reference Format:

Lorenzo Porcaro, Valeria Mirabella, Emilia Gómez, and Tiziana Catarci. 2026. Surfacing Problematic Recommender System Behaviors Affecting Music Discoverability: A Think-Aloud Protocol. In *Proceedings of the 2026*

\*The opinions expressed are those of the author(s) only and should not be considered as representative of the European Commission's official position.



This work is licensed under a Creative Commons Attribution 4.0 International License. CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/2026/04

<https://doi.org/10.1145/3772318.3791406>

CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3772318.3791406>

## 1 INTRODUCTION

Music streaming platforms are nowadays the dominant mode of online music consumption worldwide, with over 800 million subscribers by the end of 2024, generating approximately US\$20.4 billion and accounting for 69% of recorded music revenues [47, 94]. Central to the user experience on these platforms are recommender systems, which algorithmically curate content, suggest new artists, and personalize listening journeys [39]. Designed to enhance satisfaction and engagement, their design, implementation, and widespread use have brought out a series of problematic behaviors<sup>1</sup> that are today at the center of public scrutiny by scholars, policymakers, civil society, and practitioners [8]. These behaviors are closely tied to how recommender systems technically filter, rank, and optimize content exposure, with direct consequences for which music becomes discoverable to listeners.

According to the recent review by Hesmondhalgh and colleagues on the impact of algorithmically driven recommendation systems on music consumption and production [44], two central issues are: i) *popularity bias*, the tendency of recommender systems to favor already popular artists, thereby amplifying their success and limiting opportunities for emerging or independent artists; and ii) *demographic bias*, the potential for recommendations to privilege artists from certain categories or nationalities while disadvantaging others. Closely related in the literature is the concept of *fairness*, which in computer science is often framed as the mitigation of “unwanted bias” through technical approaches, whereas critical research traditions approach fairness through the lens of justice and equality. In addition, concerns around *diversity*, linked to fears that recommender systems may homogenize musical tastes and fragment consumption patterns, and the lack of *transparency* and *oversight*, frequently described through the “black box” metaphor, are also widely discussed among scholars.

Despite growing concern about such potential harms of recommender systems in the music sector, little is known about how

<sup>1</sup>In line with Bandy [5], we adopt the term “problematic,” conceiving a behavior as problematic when it causes harm (or potential harm).

end-users perceive and critically interpret these problematic behaviors. Most research on music recommendation has focused on technical evaluation [83], often overlooking the lived experiences and folk theories that listeners develop through ongoing interaction with these systems [19, 88]. Building on recent HCI research on everyday and user-driven algorithmic auditing [25, 86], as well as scholarship on algorithmic imaginaries and experiences [4, 11], this study investigates how problematic recommender system behaviors surface in daily use of music streaming platforms and how they relate to *music discoverability*. We define discoverability as the quality of being able to be found (e.g., an artist, track, or release) [12], a notion increasingly emphasized in debates on the impact of recommender systems on cultural and creative sectors [81]. Accordingly, we are guided by the following research question:

**RQ. How do end-users identify and make sense of problematic recommender system behaviors that affect music discoverability on music streaming platforms?**

Our study design builds on the seminal work of DeVos and colleagues on user-driven algorithm auditing [25], while diverging in several key aspects. First, the scenario we investigate differs from DeVos’s focus on image search. Harmful algorithmic behaviors are most commonly documented in search engines, with first investigations starting in the early 2000s [48], whereas they are less systematically studied in (music) recommender systems. In addition, the music sector operates under market logics shaped by a long-standing industry that predates streaming platforms, where what may be considered harmful often intersects with opaque promotional practices and platform–industry alignments.<sup>2</sup> Finally, the music sector is subject to a complex and evolving regulatory framework that is adapting to advancements in Artificial Intelligence (AI), autonomous decision-making, and other algorithmic systems. This, combined with limited oversight and literature on documented harms, creates a state of legal uncertainty and tension where boundaries around what is acceptable versus harmful remain blurred. Our work aims to bring greater clarity to these boundaries and contribute evidence from the underexplored domain of music recommendation.

This paper makes two main contributions to HCI: i) *conceptual*, proposing building blocks for participatory auditing of music recommender systems, outlining how listener insights can inform more accountable and equitable approaches to music discoverability; and ii) *empirical*, providing evidence of how digital-native listeners encounter and interpret problematic recommender behaviors, including the reinforcement of societal biases, the influence of commercial imperatives, and the feeling of confinement in discovery. The remainder of the paper unfolds as follows: §2 and §3 position our analysis of problematic music recommender behaviors within the broader scholarly and policy landscape; §4 outlines our methodological approach, including participant recruitment, task design, and analytical procedures; §5 reports the outcomes of the thematic analysis, which are further discussed and contextualized in §6; finally, §7 concludes the paper by summarizing the main contributions and outlining directions for future research.

<sup>2</sup>For an overview of the complex network of investments, ownerships, and strategic acquisitions shaping music and tech, see Cherie Hu, *Music Tech Ownership Ouroboros*, 2025 Edition [46].

## 2 BACKGROUND

The next sections situate our study within the broader scholarly and policy context that informs our understanding of music recommender systems and their problematic behaviors. We begin by tracing the evolution of discoverability (§2.1), from its origins in interface design to its current interpretation as a relational construct linking institutions, industry, and audiences. We then turn to music discovery in streaming platforms (§2.2), outlining how platform design reshapes listening practices and users’ attitudes toward finding new music. We subsequently introduce key mechanisms of algorithmic recommendation (§2.3) that structure content exposure on these platforms.

### 2.1 Content Discoverability

The notion of discoverability was first articulated in Norman’s seminal work [69], defined as the extent to which users can infer possible actions and how to perform them by simply looking at an interface. While central in HCI, this design-grounded conceptualization slightly diverges from the strand of work on *content discoverability*, where, from a cultural policy perspective, discoverability is oftentimes framed as a relational construct connecting institutions, industry, content, and audiences:

“Discoverability weaves its way between audience and content in a set of complex interactions consisting of marketing initiatives and strategies, but also public policies, commercial dealings, rapidly evolving business models, innovative technologies, and changing consumer habits.” [12]

Mazzoli [58] reinforces this view, arguing that discoverability “[...] concerns how different actors can play a gatekeeper role in the circulation of content online and exert more or less control over a user’s journey to content, ultimately influencing what content is deemed worthy—economically, culturally, or socially—to the final users.”

Historically, an early attempt to formalize content discoverability emerged at the 2016 *Discoverability Summit* [13], addressing the transition from broadcast to online platforms and the role of algorithmic systems [60]. Since then, discoverability has become an increasingly prominent concern in cultural policy. Examples include the Franco-Quebec mission on discoverability [97], the report by the Coalition for the Diversity of Cultural Expressions [18], the work by Mazzoli and Tambini for the Council of Europe on the discoverability of public-interest content online [59], and the UK’s Ofcom analysis of discoverability in public service broadcasting and local TV [73]. Most recently, the European Commission’s study on cultural diversity and discoverability online [32] confirms the ongoing institutional relevance of the concept.

While cultural and digital policy perspectives rightly stress distributed agency and control, our approach aligns with Johnson and colleagues’ audience-centered work on UK television [49]. There, discoverability is framed as an audience practice shaped by technological affordances, default platform behaviors, algorithmic literacy and imaginary, and ongoing negotiations with technologies, services, and content. Such framing resonates with our findings, and building on this understanding of content discoverability, we now

turn to its specific manifestations within music streaming platforms and its relation with listeners' ability and effort to discover music.

## 2.2 Discovery in Music Streaming Platforms

With the rise of streaming services, researchers began examining whether these novel modes of accessing music effectively enhanced discovery. Leong and Wright's qualitative study [53] found early evidence that streaming services offered greater opportunities for musical exploration. Aguiar's study on Deezer [1] and Datta and colleagues' study on Spotify [22] provide empirical evidence that streaming platforms became key channels for encountering new music, partly due to lower search costs and free access. Importantly, when these studies were conducted (2011–2014), streaming services were offering, for the first time, legal access to unlimited music, reframing discoverability as actively shaped by online platforms.

This interest in user attitudes and behaviors toward discovery is also reflected in research conducted by the platforms themselves. At Spotify, a team led by Garcia-Gathright conducted a mixed-method study examining user goals and satisfaction with discovery-oriented recommendations [40]. Interviews with Spotify users identified four common goals: (i) background listening of new music, (ii) listening for later immersion, (iii) finding music for future listening, and (iv) active engagement with new music. Spotify researchers have also revealed that users often discover artists at the intersections or peripheries of genre clusters [77], a result confirmed by researchers from Deezer [64].

Together, these works provide clues to the multifaceted nature of music discovery in online spaces, shaped by diverse user behaviors, goals, attitudes, and affective responses [71]. They also show the potential of recommender systems to support exploration when explicitly designed to do so, as further proven by recent longitudinal user studies by Liang and Willemsen [56] and Porcaro and colleagues [79]. While this line of research has mostly focused on analyzing music discovery practices *ex post*, our study takes a different direction by employing the Think-Aloud Protocol to capture users' real-time experiences.

## 2.3 Mechanisms of Algorithmic Recommendation

At a high level, contemporary recommender systems rely on multi-stage pipelines that combine large-scale data processing with machine learning models to curate and rank content for users. Typically, these systems separate candidate generation from ranking [20]. In the candidate generation stage, a relatively small subset of tracks or artists is typically retrieved from vast catalogs using collaborative signals (e.g., co-listening patterns, user-item interactions), content-based representations (e.g., audio features, metadata, or embeddings), and contextual information such as time, device, or listening context. In the subsequent ranking stage, candidates are ordered according to predicted user engagement or relevance, often optimizing objectives such as listening duration, skips, saves, or retention.

Consequently, discoverability is largely determined by how content is filtered and prioritized [81]. Collaborative filtering approaches, for instance, may amplify feedback loops whereby already popular artists receive disproportionate exposure, thereby reinforcing

popularity bias over time [14]. Similarly, optimization strategies focused on engagement metrics tend to favor familiar or commercially successful tracks [61], limiting opportunities for emerging, niche, or less-promoted artists. A related challenge is the trade-off between exploitation, which maximizes immediate relevance, and exploration, which introduces novelty to support long-term discovery [17]. Yet, exploration may often be constrained by business objectives [80], contributing to repetitive recommendations and a sense of confinement in musical discovery.

Importantly, these behaviors may arise from deliberate design choices around ranking objectives and data availability [66]. For auditing purposes, this matters because problematic behaviors are embedded in system structure and become apparent primarily through their effects on repeated user interactions. Understanding these mechanisms helps explain how users may indirectly perceive algorithmic choices through patterns of exposure, repetition, and omission, without requiring access to internal models or data.

## 3 RELATED WORK

Next, we discuss recent scholarship on everyday and end-user algorithmic auditing (§3.1), which highlights how ordinary users can surface algorithmic harms through situated interactions. Afterwards, we introduce the Think-Aloud Protocol (§3.2), the core methodology of this study, and review its prior applications in Music Information Research. Finally, we review existing research on perceptions of problematic behaviors in music recommender systems (§3.3).

### 3.1 Everyday and End-user Algorithmic Audits

While traditional auditing approaches often rely on domain experts [62], recent research from the CHI and CSCW communities emphasizes the potential of end-user auditing, where everyday users actively scrutinize algorithmic behavior to surface biases and harms that may otherwise remain hidden.

Shen and colleagues [86] demonstrate how end-users, through routine interactions, can reveal problematic algorithmic behaviors. Building on this, the work led by DeVos [25] examines the diverse strategies users adopt when they suspect algorithmic issues. Lam and colleagues [50] further extend this perspective by proposing a system that empowers communities to coordinate large-scale investigations of algorithmic harms, highlighting the potential of collective action. Moreover, Li and colleagues [55] analyze collaboration and division of labor in user-led audits, underscoring the social and organizational dynamics that shape how users collectively surface harms. At the same time, the work led by Deng [24] explores the challenges of embedding user-engaged auditing into industry practice, exposing the tension between theoretical approaches and the practical realities of making algorithms transparent and auditable to non-expert users.

Everyday and end-user algorithmic auditing offers several benefits. First, it allows users to surface unexpected or situated issues that may remain invisible to formal audits, drawing on their lived and contextual expertise to reveal how algorithmic behaviors affect different communities in practice [86]. In addition, by engaging diverse users, these audits can expose both converging and diverging perspectives, helping developers understand the trade-offs and

priorities embedded in system design [50]. Moreover, user-driven audits can foster early detection of harmful behaviors and create continuous accountability loops between users and developers [24]. However, there are also important limitations. End-user audits depend on users' motivation, access, and technical literacy, which can lead to self-selection biases and underrepresentation of marginalized voices [25]. Without company cooperation, data access and transparency may remain limited, and as non-technical participants, users may misinterpret algorithmic outputs or raise spurious claims. Finally, while user-led audits can highlight harms, they cannot by themselves resolve structural issues inherent to a system [86].

Collectively, these studies demonstrate that users, even without technical expertise, can identify and interpret problematic algorithmic behaviors through everyday engagement. Our research builds on this line of inquiry by focusing specifically on how end-users detect potentially problematic recommender system behaviors affecting music discoverability on streaming platforms.

### 3.2 Think-Aloud Protocol in Music Information Research

The Think-Aloud Protocol (TAP) is a qualitative method in which participants verbalize their thoughts, feelings, and actions while performing a task, providing a window into otherwise unobservable cognitive processes [30]. By externalizing mental states, TAP captures decision-making pathways, reasoning, and internal models, generating rich qualitative data [54]. Despite its strengths, TAP has seen limited application in Music Information Research (MIR), which we summarize as follows.

Lee and Price [51, 52] make use of TAP to study user experiences on commercial music services. In their studies, participants verbalized their interactions with known-item searches, playlists, and recommendations, revealing that listeners often relied on multiple platforms and were forgiving of service shortcomings. Siles and colleagues [87, 88] adapted TAP to investigate how Costa Rican Spotify users may appropriate the platform. Their studies offer insights into folk theories, playlist affective reasoning, and algorithmic personification, discussing algorithmic agency in a Global South context. Similarly, Cole [19] examined algorithmic knowledge and reflexivity in a music streaming context, finding that user understandings often resemble folk theories grounded in lived experience. Instead, Petridis and colleagues [77] applied TAP with expert music curators, who verbalized their exploration and decision-making processes while engaging with their preferred tools for music information seeking.

Methodologically closest to the approach in this study, de Assunção and Zaina [23] evaluated music discovery on Deezer and Spotify using five guided TAP tasks focused on satisfaction, cognitive load, and alignment with user preferences. Their findings revealed that while platforms performed similarly, user satisfaction and perceived control improved as tasks progressed, emphasizing the importance of multi-dimensional user experience beyond algorithmic accuracy. Instead, the study most closely aligned conceptually with our work, by Dinnissen and colleagues [28], employs TAP to investigate whether giving users insights into the fairness and diversity of recommendations can promote fairer music selections. Overall, these studies show TAP's value in uncovering users'

engagement with music recommender systems and streaming platforms. Building on this foundation, our study applies TAP to explore end-user algorithmic auditing, specifically targeting perceptions of problematic recommender system behaviors.

### 3.3 Perceptions of Problematic Music Recommender System Behaviors

The growing scholarly interest in problematic behaviors connected to algorithmically driven music recommendation is reflected in several recent literature reviews, including the aforementioned work by Hesmondhalgh and colleague [44], Dinnissen and Bauer's work on fairness [26], Ospitia-Medina and colleagues' review of biases in music recommender systems [74], and the more recent study by Carnovalini and colleagues on popularity bias [14]. While we refer the reader to these works for a comprehensive overview of the various approaches addressing these issues, in what follows we concentrate on the strand of literature that examines how people perceive such problematic behaviors.

Ferraro and colleagues [36] were among the first to investigate what is considered fair from the artists' perspective, interviewing a group of Spanish-speaking musicians. Their findings highlight that gender balance, promotion of the long tail, as well as control and transparency, are all perceived as essential elements for achieving fairness in recommender systems. Building on this work, Dinnissen and Bauer [27] extended the analysis to Dutch artists, confirming the earlier findings within a different cultural context. Still focusing on fairness, Htun and colleagues [45] conducted a user study on group music recommender systems, showing that perceptions of fairness may be shaped by individual personality traits. From a related but distinct perspective, Porcaro and colleagues [78] examined how listeners assess diversity in music recommendation, showing that various characteristics of tracks and artists influence perceived diversity, also exploring the relationship between computational diversity metrics and user perceptions.

Moreover, Ferwerda and colleagues [37] focused on popularity bias, investigating whether differences in algorithmic performance influence users' experiences. What emerged is that their study participants did not perceive any significant difference in the popularity of recommended songs or in the fairness of the recommendations. Ungruh and colleagues [98], instead, examined approaches to mitigate popularity bias, providing an in-depth analysis of their impact on user perception. Lastly, Dinnissen and colleagues [28], still addressing the intersection of popularity bias and fairness, investigated how granting users greater control over their recommendations could affect their decisions, revealing that users consistently prioritized their musical preferences over fairness-related values, leading to discrepancies between their expressed support for equitable treatment and their actual choices.

Even though this strand of literature provides valuable insights into the perception of problematic behaviors, through the use of TAP we have been able to capture users' nuanced, real-time perceptions as they navigated recommendations, offering fine-grained insights into how these problematic behaviors manifest and are recognized, a gap that has not yet been fully addressed in existing research.

## 4 METHODS

We conducted qualitative interviews with 20 Italian young adults and digital-native listeners, applying a Think-Aloud Protocol (TAP) with discovery-oriented tasks to capture how participants navigated music streaming platforms and reflected on the recommendations they encountered. While TAP is a well-established method in HCI [7], and it has already been applied to algorithmic auditing [25], it has seen limited application in MIR. Here, we use it to bridge these domains, demonstrating its potential for surfacing problematic recommender system behaviors and supporting end-user auditing practices in the music sector.

### 4.1 Positionality Statement

Our investigation is shaped by the diverse professional and academic backgrounds of our research team, all based at public institutions in Southern Europe. We recognize that our shared geographical and institutional context encourages a critical lens influenced by recent EU policy debates on algorithmic transparency and auditing. This may differ from perspectives prevalent in other regions and may, at times, narrow our focus toward EU-centric framings or solutions. Moreover, as active users of music streaming platforms, we are personally invested in music discoverability. While this familiarity provides practical insight into common platform affordances, it also introduces biases linked to our own usage patterns and expectations. Our positionality further reflects the specific expertise of the authors, combining knowledge in music recommender systems and information retrieval with a strong focus on HCI and AI ethics. These orientations can create internal tensions, which we address through ongoing reflexive dialogue. Overall, our collaborative approach leverages these varied, and occasionally conflicting, perspectives to enrich our analysis. At the same time, we acknowledge that our shared European regulatory context and disciplinary backgrounds shape our interpretation of the data, and that alternative readings may emerge from other standpoints.

### 4.2 Recruitment and Informed Consent

The Ethics Committee for Transdisciplinary Research of Sapienza University of Rome approved the study design, confirming compliance with relevant data protection, legal, and ethical standards.<sup>3</sup> A small-scale pilot study was conducted in April 2025 to refine the task definition and to test the data collection process. The main study ran from May to June 2025. Participants were recruited via *Prolific*, an online recruitment service, and compensated at £8.00/hour, consistent with the platform’s minimum pay guidelines. They were informed of the voluntary nature of their participation, their right to withdraw at any time, and their rights regarding access, rectification, and deletion of their data. Each participant received an information sheet detailing the research objectives, methodology, risks, and benefits, and provided informed consent prior to participation.

### 4.3 Prescreening

Participant prescreening followed a two-step process: first, applying predetermined criteria available on the recruiting platform, and

second, administering a custom questionnaire for which completion took approximately 10 minutes. All participants were compensated regardless of eligibility.

In the first step, participants were selected based on age (18–25), nationality and country of residence (Italy), and fluency in Italian. We ensured a balanced distribution by sex assigned at birth but did not apply other demographic filters. The age restriction targeted young adult members of Generation Z because it represents the first generation to have grown up with algorithmically mediated music discovery as a normative listening practice, primarily through streaming platforms and social media [9, 92]. Focusing on a single generational cohort minimized variability in technological familiarity and listening habits, allowing for a more controlled exploration of how users make sense of recommender system behaviors.

Moreover, restricting recruitment to Italian participants ensured cultural homogeneity in language and music industry context, which facilitated richer qualitative interpretation. This decision was further informed by evidence of the specificity of the Italian music industry market, briefly presented in Appendix A. This sampling strategy was thus designed not for generalization, but to enable an in-depth, culturally grounded analysis of a group for whom algorithmic music discovery is most pervasive.

Participants meeting these initial criteria completed an online questionnaire divided into five sections: (i) confirmation of demographic information; (ii) preferred platforms and devices for music listening and discovery; (iii) the Active Engagement subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI) [65]; (iv) the Algorithmic Media Content Awareness (AMCA) scale [101]; and (v) self-assessment of their knowledge of the algorithmic mechanisms behind music recommendations. The Active Engagement subscale of the Gold-MSI measures how much time, effort, and resources individuals devote to music-related activities such as attending concerts or exploring new music, reflecting a person’s overall involvement and investment in music. Instead, the AMCA scale is a self-report measure designed to assess users’ awareness of how algorithms shape media content on online platforms.

We made use of the demographic responses to validate the *Prolific* prescreening, while thanks to Section (ii) we verified two additional inclusion criteria: participants had to (a) select “smartphone” as their preferred music listening device and (b) use an individual account on their chosen streaming platform. The first criterion reduced variability from interface differences between desktop and mobile apps, while the second ensured that algorithmic recommendations reflected only the participant’s own activity. No restrictions were imposed on the choice of streaming platform. The remaining questionnaire sections (i.e., Gold-MSI, AMCA scale, and self-assessed knowledge) served a descriptive purpose, providing a richer profile of participants rather than testing relationships with the qualitative findings. These measures helped contextualize the sample in terms of musical sophistication, algorithmic awareness and knowledge, as detailed in the next section. A fac-simile of the pre-screening questionnaire is provided in the supplementary material.

<sup>3</sup>Protocol ID: CERT\_194BBD0C499. Digital copies of the submitted documentation, including the ethics and data protection certificates, are available upon request.

#### 4.4 Interview Settings and Think-Aloud Protocol

Interviews were conducted remotely via Zoom in Italian and lasted approximately 30-45 minutes. Each session began with an introduction outlining the study's scope and the TAP. Participants were instructed to verbalize continuously everything they were doing, seeing, and noticing while performing the tasks, with instructions deliberately limited but precise [7]. Before starting, participants were reminded of their rights as outlined in the informed consent form, and recording began once consent was given. We employed a *concurrent* TAP to capture user performance in realistic usage contexts, following established HCI guidelines [3]. Each interview comprised six think-aloud tasks and three semi-structured questions placed before, during, and after the tasks, detailed in Table 1. In addition, the slides with the original task formulation in Italian are provided in the supplementary material.

The opening question asked participants which artists they expected to encounter when opening their preferred app, establishing a baseline for their assumptions about recommender logic. The first task served as a warm-up, where participants used the app as they normally would. Feedback was provided only if participants were silent, vague, or omitted mentioning visible artists. The next three tasks focused on identifying artists who appeared (i) prominently, (ii) as novel, and (iii) as diverse from the participant's typical listening. These tasks were designed to elicit reflections on three aspects: the visibility of highly exposed artists, the discovery of unfamiliar ones, and the characterization of artists outside personal taste. Task 1 probed drivers of overexposure, while the novelty and diversity tasks connected to beyond-accuracy evaluation criteria recognized in recommender systems research [15].

After Task 3, an interlude encouraged participants to reflect on the broader role of recommender systems, particularly how artist exposure relates to user preferences and behaviors. To prevent funneling participants' responses in subsequent tasks, we refrained from introducing examples of problematic behaviors from the literature. Instead, participants were prompted to critically consider recommender influence before moving to Task 4, which focused on artists receiving less visibility, mirroring the logic of Task 1. The final tasks served as a wrap-up, with no specific instructions other than to continue using the app, allowing us to observe how earlier reflections influenced behavior. At the conclusion of the TAP, participants were presented with the study's broader goals and examples of media coverage on issues such as gender bias, popularity bias, and diversity in music recommendations. They were invited to share opinions, relate these themes to their own TAP experiences, and connect them to findings from the literature. Recordings were then stopped, and participants could ask questions or provide informal feedback. Notes were taken to capture participants' impressions and interest in the broader research agenda.

Although the task sequence was standardized, participants' app usage behaviors varied considerably. Some adopted an exploratory approach, navigating recommendations, playlists, or artist pages, while others followed habitual routines, such as directly accessing a personal playlist or library. When prompted to "keep using the app" participants typically continued in line with their usual listening habits rather than artificially exploring new features. This

heterogeneity was deliberately preserved to capture authentic interaction styles and provided valuable contextual information for interpreting their reflections during the think-aloud process.

In conclusion, the rationale for this protocol was to gradually equip participants to identify problematic behaviors in recommender systems rather than directing them toward pre-identified issues. This approach reflected the specificity of the study context and sought to avoid constraining the emergence of unanticipated concerns, consistent with principles of everyday algorithmic auditing [86]. As shown later, this method enabled the identification of harmful cases previously unknown to the research team.

#### 4.5 Thematic Analysis

We employed a hybrid approach to thematic analysis, following the six phases outlined by Willig [100] (familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report) while combining deductive and inductive coding. We began with a deductive phase to ensure analytical alignment with existing literature on user perceptions of algorithmic bias and harm, grounding our initial interpretation in established conceptual categories.

This structured entry point facilitated comparability with prior work while providing a coherent lens for the early coding of participants' talk about recommender systems. Subsequently, we shifted to an inductive approach to allow for the emergence of novel, context-specific insights beyond the initial framework. This progression from deductive to inductive analysis helped balance theoretical guidance with empirical openness, preventing the analysis from being overly constrained by predefined categories and enabling a richer, data-driven understanding of the phenomenon.

In the deductive phase, we applied four pre-identified themes of "knowledge and beliefs" about biased and harmful algorithmic behavior, drawn from DeVos and colleagues [25]. These themes, assumed to both shape and be shaped by participants' sensemaking of algorithmic behavior, are:

- (1) **Exposure and experience:** participants' prior exposures to and experiences of recommender system behaviors, focusing on what users have observed, encountered, or felt as a direct result of interacting with music recommendations.<sup>4</sup>
- (2) **Expectations and values:** participants' anticipation of how a recommender would behave in a particular situation, forming expectations about recommender system behavior.
- (3) **Folk theories of algorithms:** participants' mention of folk theories about why recommenders behave in the ways that they do, which informed their search and sensemaking process.
- (4) **Conceptions of bias and harm:** participants' definition and identification of recommender bias and harm based on, e.g., the intentions behind the design, the potential for imposition, and the priorities of recommender systems.

<sup>4</sup>We deviate from the original coding used by DeVos and colleagues, who emphasized in this theme "Participants' prior exposures to and experiences of bias" [25]. Our decision to adjust their framework is motivated by three factors: (1) the different design of our TAP tasks, (2) the specificity of the music recommendation scenario we analyzed, and (3) the absence of a consistent body of literature and shared conceptualization of problematic behaviors in music recommendation.

**Table 1: TAP Tasks and Instructions**

Task Name	Instruction
<i>Pre-TAP</i>	Before opening the app, what kinds of artists do you expect to find? Can you think of any specific names?
Task 0 (Warm-up)	Please open the app and use it as you normally would when listening to music. Speak out loud and describe what you see on the screen. Tell us about the first artists you come across.
Task 1 (Prominent)	Keep using the app and describe the artists that seem to be shown most prominently.
Task 2 (Novel)	Keep using the app and describe if you come across any artists you didn't know before.
Task 3 (Diverse)	Keep using the app and describe if you come across artists that seem different from the ones you usually encounter.
<i>Interlude</i>	Music recommendation systems can influence which songs and artists are suggested to users. In some cases, this process may lead to outcomes that favor the visibility of certain artists over others. Users' experiences with these recommendations can vary. What is your opinion on this?
Task 4 (Visibility)	As you continue using the app, describe the artists being recommended to you, especially those who seem to have less visibility compared to others.
Task 5 (Wrap up)	As we wrap up, keep browsing and describe the artists you see now.
<i>Post-TAP</i>	Our research focuses on how algorithms, such as recommendation systems, can sometimes exhibit behaviors that are considered problematic. For example, the overrepresentation of certain artists compared to others, or of certain musical cultures over others, phenomena linked to stereotyping or discrimination. To give you an idea, in this slide we have included some examples of studies that have addressed these issues. What is your opinion on this?

These themes structured our initial coding and provided an analytical lens. We then adopted an inductive approach: within each of the four themes, coded excerpts were grouped by themes to facilitate axial coding of sub-themes specific to our context, allowing novel insights to emerge. Two authors independently conducted the coding for each interview. Throughout this process, authors iteratively discussed, refined, and reconciled codes to reach consensus. Two additional authors reviewed the coding scheme and report, providing feedback on coherence and structure. This cycle of independent coding, discussion, and negotiation enhanced reliability and minimized individual bias. The final synthesis of findings under the identified sub-themes is reported in the next section, the coding scheme is part of Appendix B (Tables 3, 4, 5, and 6), while the coded excerpts are provided in the supplementary material.

## 5 RESULTS

Our analysis of the think-aloud interviews revealed three key themes in how participants perceived and made sense of problematic recommender system behaviors: i) the reinforcement of societal biases, ii) the impact of commercial imperatives, and iii) the pervasive feeling of confinement and limited discovery. Participants articulated these issues through folk theories and personal experiences, providing a rich, user-centered perspective on these problematic behaviors. The following sections detail each of these themes, using direct participant quotes to illustrate their real-world impact on music discoverability.<sup>5</sup>

<sup>5</sup>All participant quotes were originally in Italian. Translations were first produced using an automatic translation tool and subsequently reviewed and corrected by the authors for accuracy and consistency.

### 5.1 Participants

Twenty Italian young adults meeting the inclusion criteria took part in the study (11 women, 9 men, 1 non-binary), with an average age of 23.5 years (range 18–25). Most participants hold a bachelor's degree or higher (75%), and a substantial proportion are still studying (65%).<sup>6</sup> Table 2 provides a detailed overview of participant characteristics. Regarding music streaming preferences, Spotify was the most widely preferred platform for both listening (85%) and music discovery (75%). Other platforms reported include Apple Music, Amazon Music, and YouTube Music, with YouTube primarily used for discovery. Participants also frequently mentioned social media as a discovery tool (75%), reflecting broader trends in contemporary music consumption (see Appendix A). Some participants reported using different platforms for listening and discovery, reflecting the potential complementary functions of these platforms, e.g., using TikTok or YouTube to encounter new music and Spotify for regular listening.

Scores on the Active Engagement subscale of the Gold-MSI indicate slightly higher musical sophistication among participants ( $M = 4.5 \pm 1.0$  over 7) compared to the general Italian population [82], consistent with prior findings that younger and more educated individuals tend to score higher. In terms of algorithmic awareness, participants also scored relatively high ( $M = 4.2 \pm 0.5$  over 5), though the scale has not previously been applied to Italian young adults

<sup>6</sup>According to official statistics from the Italian National Institute of Statistics, currently the population of young adults between 18 and 25 is approximately 5.3 million. Most students enroll at university right after high school, at age 19. According to data from the Ministry of University and Research, approximately 1.3 million young adults are enrolled in university, between 25% and 30% of the total population. This number shows that our group of participants is somehow highly educated if compared to the overall trends in the country.

**Table 2: Participant characteristics and scores. Gold-MSI scores are reported on a 7-point Likert scale (1–7), whereas AMCA and self-assessed knowledge (Self-Know) scores are reported on a 5-point Likert scale (1–5).**

ID	Gender	Education	Platform (Listen / Discover)	Other Platforms (Discover)	Gold-MSI	AMCA	Self-Know
P1	N	Bachelor	Spotify / Spotify	YouTube, Apple Music	5.4	3.5	3.3
P2	M	Vocational school	Spotify / Spotify	SoundCloud, Bandcamp	5.1	4.4	2.8
P3	W	High school	Spotify / Spotify	YouTube	4.6	4.5	2.5
P4	W	Bachelor	Spotify / Spotify	Twitter, Instagram	3.9	4.2	1.5
P5	W	Bachelor	Spotify / Spotify	TikTok, YouTube	2.7	4.3	2.0
P6	W	Bachelor	Spotify / Spotify	TikTok, Twitter	5.9	4.4	3.0
P7	W	Bachelor	Spotify / YouTube	Instagram	4.9	4.9	2.3
P8	M	Bachelor	Spotify / YouTube	Instagram, YouTube	2.6	2.9	3.0
P9	W	Bachelor	Apple Music / YouTube	Spotify, Instagram, Twitter	5.6	4.6	4.3
P10	W	Bachelor	Spotify / Spotify	Social Media	5.1	3.9	2.0
P11	W	Master	Spotify / Spotify	Reddit, Twitter, Tumblr	4.4	4.6	3.3
P12	W	Vocational school	Spotify / Spotify	Instagram	3.1	3.9	2.5
P13	M	Master	Apple Music / Apple Music	Instagram, YouTube, Facebook	4.4	3.9	2.0
P14	M	Vocational school	Spotify / Spotify	YouTube, SoundCloud	5.8	4.5	3.8
P15	M	Master	Spotify / Spotify	Social Media	5.1	5.0	4.3
P16	M	Vocational school	Spotify / Spotify	Instagram	4.4	4.2	2.3
P17	M	Master	Spotify / Spotify	Instagram	3.8	4.2	2.5
P18	W	Bachelor	Amazon Music / Amazon Music	Instagram, Threads	2.7	4.2	2.3
P19	M	Bachelor	Spotify / Spotify	YouTube	4.6	4.2	2.8
P20	W	Master	Spotify / Spotify	Social Media	5.1	4.6	2.3

or music streaming platforms, limiting direct comparisons with the original findings [101]. Interestingly, despite apparent awareness of algorithms' roles in music recommendation, participants' self-assessed knowledge of algorithmic mechanisms was low ( $M = 2.7 \pm 0.8$  over 5). Previous work suggests that algorithmic literacy depends on both awareness and knowledge [72], and although participants reported higher awareness than factual knowledge, the combination of these factors still appears to partly support their ability to interpret the algorithmic mechanisms shaping their music listening and discovery experiences, as discussed below.

## 5.2 Reinforcement of Societal Biases: from Gender to Genre

Several participants ( $n=14$ ) consistently perceived that music recommender systems may reinforce and amplify societal biases across gender, ethnicity, language, and their intersections.

**5.2.1 Intersection of Societal Biases.** Gender bias was a recurring concern, with several participants feeling that male artists were favored, often in combination with genre-related dynamics. As P4 explained: “[...] when you look for recommendations in a genre like Italian male rap, the suggestions tend to be much more accurate, more curated, broader, and above all more relevant to the genre you are interested in, compared to, for example, female rap.” However, this perception was not universal: some described their feeds as balanced, including P1, the only participant identifying as non-binary, who noted: “I perceive it as fairly balanced, I get recommendations

for male and female artists, as well as for people who have come out as non-binary.”

Biases were also reported with respect to ethnicity and race. For instance, P1 observed: “It’s rare for Black people to come out [...] there are many Caucasian people, occasionally Asian, but African American or North African almost never.” Similarly, P5 stated: “I have very few listens from what we might call minority categories, or even, for example, artists from the Black community.” Bias was also linked to artists’ origin, with P12 remarking: “They’re always more or less the same, either some very famous American artists, or, depending on our location I suppose, the Italian ones.” Even within local contexts, language could intersect with visibility, disadvantaging artists performing in a dialect. As P4 put it, mentioning an artist who sings in Neapolitan dialect: “I do not think he has as much visibility as other Italian artists, because he doesn’t sing in Italian but in dialect, and his music belongs to a genre that is not typically considered viral.”

**5.2.2 Perceived Causes and Mitigation Strategies.** Strikingly, many participants accepted these reinforced societal biases as a “natural consequence,” attributing them to three main factors: (i) industry-level imbalances (“The fact that the artists are mostly male rather than female is [...] kind of natural, in the sense that, from the start, there are simply more male artists”, P7); (ii) algorithm design (“I think there is a bias in these algorithms that comes from the very nature of the algorithm itself”, P15); and (iii) personal listening habits (“I tend to listen to male artists more often [...] so I assumed the imbalance was due to the way I used the app”, P17).



Some participants described individual strategies to counteract these biases, such as deliberately seeking out underrepresented artists. P20, for example, explained: *"I made a conscious effort to listen to more female artists in the genres I usually follow, because they did not come up very often."* Others, however, expressed resignation: *"To change the way the algorithm suggests a new song, you would have to change the stigmas of society, and as we all know, that's an almost impossible task"* (P4). P20 also discussed platform-led interventions, such as gender-based playlists, however, with skepticism: *"I get playlists like emerging female artists in a certain genre, but I tend to ignore them [...] they include songs made by women, yes, but not in the genre I'm looking for, so I just skip them."*

### 5.3 Commercial Imperatives: Virality, Strategic Placement, and Forced Exposure

More than half of the participants (n=12) consistently reported that commercial interests and market dynamics strongly influence music recommender systems, leading to the strategic promotion of certain artists and a sense of forced exposure to popular content.

**5.3.1 Popularity, Virality and Spillover Effects.** The push toward well-known artists, i.e., the popularity bias, was frequently noted and often negatively received. For example, P11 remarked: *"There's often a strong push for certain artists, either because they've recently gone viral or because there's a major label backing their promotion. It happens a lot, and it can be a bit annoying."* Emerging artists were perceived as particularly disadvantaged, as P3 stated: *"It's more likely that you'll be shown songs that are already popular [...] and that's a problem, because for an artist trying to emerge, it becomes much more complicated."* P17 emphasized that the opacity of recommender systems constituted a limiting factor: *"As an emerging artist, it's really hard to break into dynamics where your chances of success depend on an algorithm you don't even understand."*

Participants distinguished the influence of virality, rapid and widespread dissemination, from popularity, which reflects sustained recognition over time. P10 observed: *"Since it's something that's viral at the moment, they might be trying to push the artist more."* Also P4 described virality as a strategic focus: *"I think what Spotify wants to do is focus on going viral rather than helping you discover new artists."* However, P1 noted that virality can sometimes benefit niche artists: *"Those who maybe have more frequent listeners, or who have a specific track that has gone viral, still emerge."*

Social media platforms, especially TikTok, were seen as central to virality, as P2 explained: *"By now it's like the major player in the attention market."* P10 added: *"I know they have songs going viral on TikTok, so I understand that they might be taking advantage of that moment. The frequent use of social media actually helps certain songs become popular again."* Instagram was also mentioned, though less prominently.

**5.3.2 From Strategic Placement to Forced Exposure.** Participants identified two main drivers of these dynamics: retaining users on the platform and maximizing revenue. P19 emphasized engagement: *"What really matters to music platforms is that people spend as much time as possible on the app, and so having an algorithm that works means giving someone something they're very likely to enjoy."* P15 framed this as a legitimate business objective: *"It's something music*

*platforms do on purpose to keep users on their platform more and more, which, as a company, I can understand."* Others highlighted the profit motive, as P7: *"I think the goal is basically to push whoever brings in the most money, to push what can sell more."* P19 nonetheless contextualized this behavior as not specific to online spaces: *"It could also happen in a record store [...] the store preferred buying and selling CDs from well-known artists rather than more niche ones."*

Participants developed folk theories about how commercial incentives shape artist recommendations. P12, for instance, wondered about sponsorship: *"I rather wondered whether there might be a kind of sponsorship dynamic, like making an artist appear more often in other people's playlists."* Instead, P3 questioned payment structures: *"I don't really know how it works for an artist to publish on Spotify, and whether they have to pay to stay visible."* P7 was more certain: *"There surely is... there's an algorithm that pushes certain artists, maybe through payments."* At the same time, P5 noted that recommendations are not entirely independent of personal taste: *"I don't think some artists get pushed independently of the user's taste."*

Participants also discussed collaborative and placement dynamics. For example, P2 highlighted next-track suggestions: *"Being the next suggested artist after a song is a really strong placement, it's a great spot with huge benefits."* P11 observed that platforms mix popular and niche artists: *"It usually shows a big artist first and then a smaller one"*, interpreted by P20 as a risk-averse strategy: *"Recommending slightly smaller artists can indeed be riskier."*

Awareness of these mechanisms did not make participants more accepting of the recommendations, especially when they felt forced to listen to artists. P11 described the friction: *"It bothers me, especially when they recommend a genre I'm not interested in and still try to push me to listen to an artist I don't care about."* Specific cases where repeated exposure prompted social media discussions, such as Sabrina Carpenter's "Espresso", were reported by P11 and P17: *"When a playlist or an album I had chosen ended, Spotify would start playing its recommended songs, and it was always 'Espresso' by Sabrina Carpenter [...] I even saw on Twitter that other people had the same experience"* (P11).

Participants often felt powerless to change these dynamics. For example, P2 reflected: *"It's true that some artists can have wider visibility, but it's not a dynamic you can't escape or that leaves no alternatives."* P1 highlighted the limits of feedback mechanisms: *"In some cases I even hit 'dislike' because I'm not interested, [but] they still keep being recommended."* Others perceived the issue as minor, as P17: *"Having access to artists I don't know, even if they're promoted, isn't a major issue."*

### 5.4 Confinement and Limited Discovery: the Role of Preferences and Historical Data

Participants (n=14) frequently described a sense of confinement in music recommendation, which they attributed to the recommenders' heavy reliance on their personal preferences. They referred to this phenomenon as "own niche" (P11), "echo chamber" or "comfort zone" (P6), and "own bubble" (P7), employing terms oftentimes (mis)used in the public and scientific debate.

**5.4.1 The Harms of Algorithmic Confinement.** Several participants perceived the algorithms as limited in scope, overly specific, or stuck in particular listening patterns. P11 noted: *"When it's a genre*

*I know a bit better, I realize how limited the algorithm is, because it keeps trying to push you towards the more famous artists.”* P15 explained: *“The problem is when the algorithm becomes so specific that it no longer lets you discover new music.”* Mis-targeting was also observed by P11, who remarked: *“It recommends a lot of music that I consider very much for teenagers [...] I’ve somewhat passed adolescence and yet it stubbornly continues to recommend these artists to me.”* Others highlighted repetition of overly popular artists (P18) or persistent focus on already familiar ones (P20). This pattern restricted participants’ ability to explore, particularly for those seeking to broaden their musical taste. As P19 put it: *“Very often I get recommended the same things over and over, so it becomes a bit harder to broaden what I listen to.”* P6 generalized the phenomenon beyond streaming, noting: *“As on all social platforms [...] the same things keep resonating almost obsessively.”*

These phenomena may contribute, according to several participants, to harms at both the individual and societal levels. At the individual level, confined recommendations risk driving withdrawal and closure: *“I think there is a kind of overall closure, and that’s a danger, because we end up closing ourselves off”* (P3). Another participant described becoming accustomed to the comfort of one’s bubble, finding it satisfying, yet ultimately limiting discovery: *“I get used to the bubble I’m in, and I’m okay with it, because in the end, most of the time, it hits the mark, [though] it would really please me to be exposed to different things, but it never happens”* (P7). At the societal and cultural level, P13 reflected on how such confinement might contribute to the standardization of music: *“They tend to suggest the same music genres over and over, always the same ones, even down to the choruses [...] this can lead to a sort of flattening.”*

Tellingly, breaking these loops was described as difficult. For example, P20 noted: *“I’ve had quite a hard time in the past breaking out of these loops [...] it becomes harder to realize when it has started circling back on itself and to interrupt the loop.”* Still, some participants sought collective recognition as a first step toward awareness: *“It happens very often that the music it recommends to me is always the same [...] I was curious to see if anyone else had gone through it, and indeed other people were complaining about the same thing”* (P19).

**5.4.2 Preferences, Historical Data and Other Signals.** Confinement was closely linked to the use of historical listening data. Participants noted that recommendations often reflected their past behaviors at the artist (P1), album (P11), or genre level (P13). P4 observed: *“Everything I currently have on my homepage is based on what I listen to [...] so they kind of amplify the genre I already listen to rather than helping me discover new ones.”* Some participants acknowledged a potential tradeoff, as P20: *“When I select a song to start from, it plays about ten songs I already know before it starts presenting me with new songs.”* According to P14, the timeframe for influence varied widely, from days to months or even years. Moreover, new releases were often promoted based on prior listening, as P11 explained: *“I assume they cross-referenced artists I listened to in the past [...] who have released something new, and it’s trying to get me to listen to them again.”* P20 emphasized recency as pivotal: *“Whenever a new artist releases an album, it always puts it right at the top of the things to listen to.”*

By relying on previous listening and exploiting listeners’ tastes, algorithms essentially play it safe, as P3 commented: *“The algorithm suggests music based on criteria like how much it thinks I’ll like what it recommends [...] probably to keep the app active, to keep me on the app, so I don’t get bored.”* This logic is not necessarily negative, as P14 noted, because it can help artists better reach their audiences: *“It’s definitely a favor to the artists from this point of view too, because they could gain an extra fan, so to speak, and expand their audience thanks to the algorithm.”* P19 also highlighted how emerging artists could benefit from algorithms: *“[...] because if it notices that someone listens to more niche artists, it’s much more likely to start recommending niche artists.”*

Beyond personal listening habits, a few participants reported that relational and contextual signals can also influence recommendations. For example, P1 hypothesized that an artist recommendation appeared because *“[...] there are some friends I follow on Spotify who are actually big fans.”* Sharing music outside the platform also seems influential, according to P3: *“I’ve noticed that the people I share music with, maybe through social media or WhatsApp, get suggested artists that I listen to.”* Finally, P6 believes that location could inform the recommendation: *“Depending on where someone is, specific recommendations are also made”,* a belief reinforced by P12’s experience of being recommended music from a local event.

Live appearances at acclaimed events, such as music festivals or TV shows, were indeed hypothesized as influential signals for recommendation. P9 referred to artists featured in the *Eurovision* song contest, a local talent show (*Amici*), and a local music festival (*Sanremo*): *“Especially when there are particular events, the app’s recommendations change a lot. So specifically, it often suggests all the artists who are participating.”* Other participants noticed how experiences in external non-music platforms can also have a role. Three participants found recommendations of the soundtrack of TV series, *Arkane* (P1) and *The Crown* (P7), and a movie, *Hunger Games* (P5). Surprisingly, P7 and P5 said they have not seen those before.

## 6 DISCUSSION

Throughout the interactions of a group of Italian young adult music listeners with streaming platforms, we identified three key areas affected by problematic recommender system behaviors: reinforcement of societal bias, commercial imperatives, and confinement in discovery. Although these phenomena are not entirely new, our study uniquely explores them through the imaginary and experiences of a digital-native generation. While these explanations remain *folk theories* rather than verifiable platform strategies, they offer valuable insight into how users make sense of opaque algorithmic systems.

Hereafter, we aim to re-contextualize these problematic behaviors within the current musical ecosystem, offering a fresh perspective on long-standing issues. We then examine the implications for music discoverability, showing how these behaviors may reshape the power dynamics between platforms, artists, and listeners. Finally, we propose building blocks for end-user recommender system auditing in the music sector, outlining a participatory approach that leverages user insights to design more equitable systems.

## 6.1 Re-contextualizing Problematic Music Recommender Systems Behaviors

The impact of algorithmically driven recommendations on music listening, and in particular questions of bias and fairness, diversity and inclusion, transparency, opacity, and oversight, has been extensively explored in the scientific literature [44]. However, much of the existing work lacks a critical lens, underestimates the complexity of interactions between people and technology, and often fails to examine these questions through the lens of what listeners actually do with recommendations, rather than merely assuming their behaviors and preferences [43].

Specifically, the popularity bias and the long-tail problem, early theorized in the music sector almost two decades ago [16], the gender bias and overexposure of male artists [31], or the platforms and industry exercise of some form of power through algorithmic curation [2], are definitely known phenomena in the digital music sector. Under this lens, at first glance, the presented results may appear not novel and obvious. Nonetheless, the identification and exploration of those through the viewpoint of a group of young adults, some of whom were not even born or newborn when streaming platforms started to be launched, provides new insights and hypotheses. In fact, while this cohort of listeners has been the subject of analyses on generational differences in the use of online music platforms [70], their involvement in algorithmic auditing practices, at least within the music sector, is absent.

**6.1.1 Intersectional Representational Harms.** Participants often described feeling that the artists promoted to them reflect wider inequalities, suggesting that listeners perceive algorithmic representation as mirroring social hierarchies rather than correcting them, a perception often shared by artists themselves [27, 36]. Under this lens, a first emerging hypothesis is that representational harms in music recommender systems should be understood through an intersectional lens [21], in line with recent proposals from the algorithmic fairness community [38, 99]. Dimensions such as gender, ethnicity, language, mainstreamness, genre, and local versus global contexts may intersect in shaping recommender outcomes. Nonetheless, this intersectional complexity may complicate listeners' assessments [78], particularly when perceptions of problematic behaviors involve both musical and artist characteristics. While these factors have often been examined in isolation [44] and under fairness assumptions [26], music recommender systems research still lacks a nuanced account of their overlaps.

**6.1.2 Commercial Manipulation and Opaque Promotion.** Several participants explicitly perceived forms of commercial manipulation, believing that viral or heavily promoted tracks were “pushed” for profit rather than preference, consistent with previous findings [28]. In this direction, our results highlight the role of platforms' commercial logics and the opaque algorithmic processes of promotion. This opaqueness can produce a sense of coercion, leaving listeners feeling trapped [84] when recommendations repeatedly highlight already dominant artists without clear justification, as illustrated by the “Espresso” case described in §5.3.2. The case has become so widespread that it attracted media attention [68], sometimes dismissed as a conspiracy theory [57].

Engagement-driven, industry-led mechanisms regulating which artists are promoted reframe the longstanding popularity bias issue: the key question is not only how recommender systems help the rich get richer, but also which dynamics establish who becomes rich in the first place. At the same time, a tension exists between users' perceptions and bias measured using computational metrics [37]. Indeed, users may fail to notice popularity bias even when it is statistically high, suggesting that participants' folk theories may be triggered by specific cases rather than the overall distribution of recommendations.

In the participants' narratives, virality plays a crucial role today, amplified by cross-platform externalities between streaming services and social media, and the 2025 Music Impact Report from TikTok confirms this: 84% of songs that entered the Billboard Global 200 in 2024 went viral on the social platform first [95]. This phenomenon is even reinforced by new integrations that allow users to transfer discoveries directly into streaming libraries (e.g., both in TikTok [96] and Instagram [93]). Yet recommender systems research has paid little attention to these *spillover* effects, even though they have already been identified as core elements that should be incorporated into the design of fair ranking models [76].

**6.1.3 Confinement and Limited Discovery.** Participants frequently expressed frustration with the repetitiveness of recommendations and their inability to “escape” familiar artists, even when they sought novelty. While issues of confinement and limited discovery in online spaces are discussed in the literature, they remain problematic due to ambiguous terminology, evident in our participants' use (and misuse) of terms such as “filter bubble” and “echo chamber”. Media studies scholars have long warned that this ambiguity can misdirect research agendas [10, 63], especially as these terms enter everyday discourse. Our findings reveal a tension between participants' perception of overexposure to tailored content, its hypothesized link to historical data and inferred preferences, and the still-limited empirical evidence of its actual manifestation, despite emerging new perspectives [85].

## 6.2 Implications for Music Discoverability

Ultimately, the results of this study suggest a need for a shift in perspective. To truly enhance music discoverability, the design of recommender systems must move beyond purely technical metrics of engagement [66] and incorporate a deeper understanding of user desires for serendipity, diversity, and genuine exploration [35].

**6.2.1 Tensions and Ambiguities in Music Discoverability.** A critical tension between the technical efficiency of recommender systems and their role in music discoverability emerges from our findings. First, discoverability appears to be influenced not only by musical aesthetics and functions but also, perhaps more fundamentally, by broader social identities, with these two dimensions remaining closely intertwined. Second, similar to the pre-streaming era, music discoverability remains strongly shaped by market forces, contradicting narratives of democratization in media industries driven by new technologies [41, 42].

Third, music discoverability occupies an ambiguous position. While recommender systems are designed to match users with content they might enjoy, our study suggests that, at least for Italian

digital native listeners, recommender systems may be perceived as hindering rather than helping music discovery. From a normative standpoint, minimizing or eliminating confinement appears desirable. Yet from an artist's standpoint, the challenge may not be to dismantle the logic entirely but rather to find pathways, supported by their audiences [6], to enter existing circles or to create new ones. From a user's perspective, breaking the confinement loop requires a trade-off. In fact, mitigating bias may reduce familiarity [98], but it is precisely this reduction that can evoke the sense of discovery participants sought.

**6.2.2 Popularity Bias, Overexposure, and Virality.** The folk theories and experiences shared by participants confirm that current recommendation models may prioritize the promotion of already popular content, leading to a state of overexposure to a limited pool of artists and genres. Not surprisingly, even within the music industry the impact of algorithmic curation is tangible and acknowledged. The following quote by Sophie Lyon, Digital Sales Manager for Italy, Spain, and Portugal at BMG, included in the 2025 report of the Federation of the Italian Music Industry (FIMI) [34], makes this strikingly clear:

"The growth in the number of streams generated by albums is a positive trend that demonstrates how Digital Service Providers have become the main point of access to music. However, the vast amount of content, combined with the central role of algorithms, can limit the discovery of new artists—with only a few managing to break through." (p. 18)

This behavior has significant implications for both artists and listeners. For artists, especially emerging or niche ones, the system's focus on mainstream content may contribute to a barrier to entry, potentially limiting their ability to reach new audiences. From the listener's perspective, this over-reliance on a hit-based culture leads to a sense of confinement, trapping them in a predictable loop of familiar sounds. This goes against the core promise of music platforms to offer a vast and diverse library for exploration.

Furthermore, the amplification of viral loops, where social media trends directly influence algorithmic recommendations, creates a discoverability model that is more reactive than proactive. Instead of guiding users to new and diverse music, the system is influenced by external popularity contests, which may not align with an individual's personal taste or desire for exploration.

### 6.3 Design Considerations for End-User Recommender Systems Audits

The insights from our study suggest that the proposals from the end-user algorithmic auditing literature may provide a foundation for developing auditing practices also for music recommender systems. Effective approaches must move beyond technical critiques to engage with the subjective and social dimensions of algorithmic harms. Indeed, for our study participants, recommender bias was not an abstract concern but a tangible element of everyday music discovery and listening. They recognized that their habits were shaped by confinement and by the promotion of certain artists for reasons sometimes beyond personal taste. While often lacking technical vocabulary, their awareness, formed through constant

interaction with these systems, offers valuable starting points for designing user-centered auditing methods.

**6.3.1 Leverage Users' Folk Theories and Experiences.** End-user auditing frameworks should build on users' folk theories and lived experiences of algorithmic behavior [86]. Participants' explanations of phenomena such as virality and social media spillovers indicate that users already construct (hypothetical) causal accounts of how recommendations emerge. Future tools could support this reasoning by enabling users to annotate their experiences and provide feedback, thereby crowdsourcing the identification of algorithmic anomalies. However, auditing tools must account for heterogeneity in how users value potentially problematic behaviors [45], as individual characteristics shape the interpretation of recommendation lists.

**6.3.2 Empower Users to Act on Insights.** Auditing tools should respond to the feelings of powerlessness and limited agency expressed by participants, who noted that altering recommendations would require changing "the stigmas of society". Such tools could empower users [50], helping them move from passive frustration to active engagement, encouraging diversification of listening practices and more targeted feedback. This sense of powerlessness is further complicated by the fact that, even when users are aware of problematic issues, they often default to familiar tastes [28], revealing a gap between expressed values and actual behavior. Auditing tools should therefore be designed to make this gap visible and actionable, rather than assuming that awareness alone leads to change.

**6.3.3 Support Collective and Participatory Practices.** Auditing should be framed as a collective practice [55], fostering dialogue among stakeholders. Participants frequently sought validation for their experiences, such as sharing frustrations about the song "Espresso" on X/Twitter. This can be seen as an example of everyday end-user auditing, where users identify problematic behavior and raise awareness through social media. At the same time, artists, like our participants, often feel sidelined by opaque commercial logics and seek greater transparency in how their music is promoted and distributed [27, 36].

Future tools might enable the comparison of recommendation patterns across users, facilitating the collective identification of problematic behaviors and amplifying advocacy for more equitable systems. Indeed, we view participatory approaches to recommender system design, development, evaluation, and auditing as a long-term challenge requiring sustained research, in line with Ekstrand and colleagues' recent proposal [29]. Our contribution outlines concrete, accessible pathways toward participatory auditing methods and tools, while acknowledging that the mechanisms underlying such participatory processes remain largely conceptual. These pathways aim to enable listeners to better understand and interrogate the recommender systems shaping their everyday music discovery, and artists to make sense of the systems shaping their discoverability.

**6.3.4 Towards Prototypes and Operationalization.** Future work will expand this line of inquiry by developing and testing low-fidelity prototypes (e.g., sketches or wireframes) to visualize and operationalize participatory auditing workflows. These prototypes will serve as a means to explore how users might document, share, and

collectively interpret their experiences with algorithmic recommendations. In addition, subsequent research will aim to translate the study's qualitative insights into concrete design guidelines for recommender systems, bridging the gap between users' mental models, how they make sense of algorithmic behavior, and the systems' actual functioning, as already explored in video-streaming contexts [67]. Doing so will help transform the interpretive knowledge emerging from users' folk theories into actionable principles for more transparent and accountable recommendation design.

## 6.4 Limitations

Our study provides important insights into how Italian young adults perceive problematic behaviors in music recommender systems, yet several limitations should be acknowledged.

**6.4.1 Sample Demographics and Cultural Context.** The demographic homogeneity of our sample is a key constraint, even if participants were intentionally recruited among Italian young adults. While this design allowed for a focused, culturally specific analysis, it limits the generalization of our findings to older populations, listeners in other regions, or those with different educational and socioeconomic backgrounds. Furthermore, the theoretical grounding of this research is primarily European and Western. Several elements, such as the policy-oriented perspectives on discoverability, the platforms participants used during the TAP, and the strand of research on end-user auditing (mostly from US scholars), contribute to circumscribing our findings and discussion within a specific regional context.

**6.4.2 Self-Reported Data and Folk Theories.** Our analysis relies heavily on self-reported data and participants' folk theories of algorithmic functioning. Although these accounts highlight meaningful user perspectives, they are not always technically accurate and may reflect a gap between perceived and actual algorithmic mechanisms. Moreover, these perspectives do not explicitly account for artists' experiences. Even though some participants identified as musicians (hobbyists or professionals), the study was designed to elicit responses from the listener's perspective. This discrepancy shapes the type of data we were able to collect and interpret. Future work will focus specifically on music artists to gain a complementary, yet essential, understanding of problematic algorithmic behaviors affecting music discoverability.

**6.4.3 Scope and Methodological Constraints.** This paper reports only the first phase of a larger multi-part study. We chose to focus exclusively on this phase in order to share our findings with the CHI and HCI community and receive feedback to guide subsequent research activities. Methodologically, while the TAP captures real-time cognitive processes and immediate user reflections, it does not reveal long-term behavioral patterns or shifts in perception.

Moreover, the TAP itself may have influenced participants' behavior. Prompting users to interact explicitly with recommender systems likely increased their awareness of potential biases in ways that differ from everyday, unprompted use, potentially affecting the naturalness of interactions. Integrating findings from subsequent phases will allow future work to provide a more comprehensive and ecologically valid account of user interactions with problematic music recommender system behaviors. In the next stage of this

research, we will complement TAP with Ecological Momentary Assessment methods to capture more naturalistic listening and discovery practices in everyday contexts.

## 7 CONCLUSION

This study set out to explore how end-users of music streaming services identify and make sense of problematic recommender system behaviors. Through a user-centered, qualitative approach using a think-aloud protocol, we have shed light on the lived experiences of Italian young adults as they navigate the complexities of algorithmic curation.

Our findings demonstrate that music listeners are not passive recipients of recommendations, rather, they actively develop folk theories and personal strategies to cope with perceived algorithmic shortcomings. The themes that emerged from our analysis underscore that perceived recommender harms are not an abstract concept for users, but a tangible and frustrating element of their everyday music discovery. These behaviors can significantly influence music discoverability, as users are increasingly confined to a narrow set of recommendations and face greater difficulty in finding new and diverse artists outside of their established listening habits.

The insights gleaned from this research have several implications for the design of future recommender systems. By understanding how users perceive and articulate these issues, designers can move toward creating systems that are more transparent, controllable, and responsive to user needs. Moving forward, platforms should consider incorporating end-user auditing tools that allow for direct feedback and annotation of algorithmic anomalies. Lastly, our study contributes to the growing body of literature advocating for an end-user approach to algorithmic auditing. We argue that the success of recommender systems should not be measured solely by engagement metrics, but also by their ability to foster a healthy, diverse, and equitable music discovery ecosystem for all listeners and artists.

## Acknowledgments

This research is part of the project *Algorithmic Auditing for Music Discoverability (AA4MD)* which has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101148443. This work is also partially supported by the HUMAINT programme (Human Behaviour and Machine Intelligence), Joint Research Centre, European Commission.

## References

- [1] Luis Aguiar. 2017. Let the music play? Free streaming and its effects on digital music consumption. *Information Economics and Policy* 41 (2017), 1–14. doi:10.1016/j.infoecopol.2017.06.002
- [2] Luis Aguiar and Joel Waldfogel. 2021. Platforms, Power, and Promotion: Evidence from Spotify Playlists. *The Journal of Industrial Economics* 69, 3 (2021), 653–691. doi:10.1111/joie.12263
- [3] Obead Alhadreti and Pam Mayhew. 2018. Rethinking Thinking Aloud: A Comparison of Three Think-Aloud Protocols. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173618
- [4] Oscar Alvarado and Annika Waern. 2018. Towards Algorithmic Experience: Initial Efforts for Social Media Contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18).

- Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173860
- [5] Jack Bandy. 2021. Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 74 (April 2021), 34 pages. doi:10.1145/3449148
  - [6] Nancy K. Baym. 2018. *Playing to the Crowd: Musicians, Audiences, and the Intimate Work of Connection*. NYU Press, New York.
  - [7] M. Ted Boren and Judith Ramey. 2000. Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication* 43, 3 (2000), 261–278. doi:10.1109/47.867942
  - [8] Georgina Born, Jeremy Morris, Fernando Diaz, and Ashton Anderson. 2021. *Artificial Intelligence, Music Recommendation, and the Curation of Culture*. Technical Report. Schwartz Reisman Institute; University of Toronto. <https://utoronto.scholaris.ca/server/api/core/bitstreams/3fee35e6-48e7-444a-8b4f-e40afadd7bae/content> Accessed: 2026-01-21.
  - [9] Graeme Bruce. 2022. *How are global consumers finding new music?* <https://yougov.co.uk/technology/articles/40255-how-are-global-consumers-finding-new-music> Accessed: 2026-01-21.
  - [10] Axel Bruns. 2019. *Are Filter Bubbles Real?* John Wiley & Sons, Hoboken, NJ.
  - [11] Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20, 1 (2017), 30–44. doi:10.1080/1369118X.2016.1154086
  - [12] Canada Media Fund (CMF). 2021. *Discoverability: Toward a Common Frame of Reference*. Technical Report. Canada Media Fund (CMF). [https://cmf-fmc.ca/wp-content/uploads/2021/01/CMF\\_-\\_Discoverability\\_Toward\\_a\\_Common\\_Frame\\_of\\_Reference\\_-\\_Final.pdf](https://cmf-fmc.ca/wp-content/uploads/2021/01/CMF_-_Discoverability_Toward_a_Common_Frame_of_Reference_-_Final.pdf) Accessed: 2026-01-21.
  - [13] Canadian Radio-television and Telecommunications Commission and the National Film Board. 2016. *Discoverability Summit: Content in the Age of Abundance*. <https://web.archive.org/web/20170505013548/http://www.discoverability.ca/> Accessed: 2026-01-21.
  - [14] Filippo Carnovallini, Antonio Rodà, and Geraint A. Wiggins. 2025. Popularity Bias in Recommender Systems: The Search for Fairness in the Long Tail. *Information* 16, 2 (Feb. 2025), 151. doi:10.3390/info16020151
  - [15] Pablo Castells, Neil Hurley, and Saúl Vargas. 2022. Novelty and Diversity in Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, 603–646. doi:10.1007/978-1-0716-2197-4\_16
  - [16] Óscar 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, Nevada) (NETFLIX '08). Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. doi:10.1145/1722149.1722154
  - [17] Minmin Chen, Yuyan Wang, Can Xu, Ya Le, Mohit Sharma, Lee Richardson, Su-Lin Wu, and Ed Chi. 2021. Values of User Exploration in Recommender Systems. In *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam, Netherlands) (RecSys '21). Association for Computing Machinery, New York, NY, USA, 85–95. doi:10.1145/3460231.3474236
  - [18] Coalition for the Diversity of Cultural Expressions (CDCE). 2020. *The Challenge of Discoverability: The Urgency to Act to Defend the Diversity of Cultural Expressions*. Technical Report. Coalition for the Diversity of Cultural Expressions (CDCE). [https://cdce-cdce.org/wp-content/uploads/2020/12/The-challenge-of-discoverability\\_CDCE.pdf](https://cdce-cdce.org/wp-content/uploads/2020/12/The-challenge-of-discoverability_CDCE.pdf) Accessed: 2026-01-21.
  - [19] Sebastian Cole. 2024. Exploring users' algorithmic knowledge and reflexivity in a music streaming context: A critical realist approach. *Convergence* 30, 6 (2024), 2075–2089. doi:10.1177/13548565241277379
  - [20] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (RecSys '16). Association for Computing Machinery, New York, NY, USA, 191–198. doi:10.1145/2959100.2959190
  - [21] Kimberlé Crenshaw. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum* 1 (1989), 139–167.
  - [22] Hannes Datta, George Knox, and Bart J. Bronnenberg. 2018. Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science* 37, 1 (2018), 5–21. doi:10.1287/mksc.2017.1051
  - [23] Willian Garcias de Assunção and Luciana Aparecida Martinez Zaina. 2022. Evaluating user experience in music discovery on Deezer and Spotify. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems* (Diamantina, Brazil) (IHC '22). Association for Computing Machinery, New York, NY, USA, Article 15, 11 pages. doi:10.1145/3554364.3560901
  - [24] Wesley Hanwen Deng, Boyuan Guo, Alicia Devrio, Hong Shen, Motahhare Eslami, and Kenneth Holstein. 2023. Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 377, 18 pages. doi:10.1145/3544548.3581026
  - [25] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 626, 19 pages. doi:10.1145/3491102.3517441
  - [26] Karlijn Dinnissen and Christine Bauer. 2022. Fairness in Music Recommender Systems: A Stakeholder-Centered Mini Review. *Frontiers in Big Data* 5 (2022), 9 pages. doi:10.3389/fdata.2022.913608
  - [27] 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. doi:10.1145/3565472.3592960
  - [28] Karlijn Dinnissen, Shah Noor Khan, Hanna Hauptmann, Eelco Herder, and Judith Masthoff. 2025. The Role of Fairness and Diversity in User Choices and Perceptions of Music Playlists. In *Proceedings of the 36th ACM Conference on Hypertext and Social Media* (HT '25). Association for Computing Machinery, New York, NY, USA, 48–57. doi:10.1145/1145/3720553.3746671
  - [29] Michael D. Ekstrand, Afsaneh Razi, Aleksandra Sarcevic, Maria Soledad Pera, Robin Burke, and Katherine Landau Wright. 2025. Recommending With, Not For: Co-Designing Recommender Systems for Social Good. *ACM Trans. Recomm. Syst.* Just Accepted (Aug. 2025), 23 pages. doi:10.1145/3759261
  - [30] K. Anders Ericsson and Herbert A. Simon. 1980. Verbal reports as data. *Psychological Review* 87, 3 (1980), 215–251. doi:10.1037/0033-295X.87.3.215
  - [31] Maria C. Eriksson and Anna Johansson. 2017. Tracking Gendered Streams. *Culture Unbound* 9, 2 (Oct. 2017), 163–183. doi:10.3384/cu.2000.1525.1792163
  - [32] European Commission. 2024. *European Commission launches study on discoverability and promotion of cultural diversity online*. <https://culture.ec.europa.eu/news/european-commission-launches-study-on-discoverability-and-promotion-of-cultural-diversity-online> Accessed: 2026-01-21.
  - [33] Federazione Industria Musicale Italiana (FIMI). 2019. *Music Listening 2019*. Technical Report. Federazione Industria Musicale Italiana (FIMI). [https://www.fimi.it/wp-content/uploads/2025/06/music\\_listening\\_2019.pdf](https://www.fimi.it/wp-content/uploads/2025/06/music_listening_2019.pdf) Accessed: 2026-01-21.
  - [34] Federazione Industria Musicale Italiana (FIMI). 2025. *Il mercato discografico italiano | Report FIMI 2025*. Technical Report. Federazione Industria Musicale Italiana (FIMI). [https://www.fimi.it/wp-content/uploads/2025/04/report\\_fimi\\_2025.pdf](https://www.fimi.it/wp-content/uploads/2025/04/report_fimi_2025.pdf) Accessed: 2026-01-21.
  - [35] Andres Ferraro, Gustavo Ferreira, Fernando Diaz, and Georgina Born. 2024. Measuring Commonality in Recommendation of Cultural Content to Strengthen Cultural Citizenship. *ACM Trans. Recomm. Syst.* 2, 1, Article 10 (March 2024), 32 pages. doi:10.1145/3643138
  - [36] 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*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 562–584.
  - [37] Bruce Ferwerda, Eveline Ingesson, Michaela Berndt, and Markus Schedl. 2023. I Don't Care How Popular You Are! Investigating Popularity Bias in Music Recommendations from a User's Perspective. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) (CHIIR '23). Association for Computing Machinery, New York, NY, USA, 357–361. doi:10.1145/3576840.3578287
  - [38] James R. Foulds, Rashidul Islam, Kamrun Naher Keya, and Shimei Pan. 2020. An Intersectional Definition of Fairness. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, Piscataway, NJ, 1918–1921. doi:10.1109/ICDE48307.2020.00203
  - [39] Sophie Freeman, Martin Gibbs, and Bjorn Nansen. 2023. Personalised But Impersonal: Listeners' Experiences of Algorithmic Curation on Music Streaming Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 412, 14 pages. doi:10.1145/3544548.3581492
  - [40] Jean Garcia-Gathright, Brian St. Thomas, Christine Hosey, Zahra Nazari, and Fernando Diaz. 2018. Understanding and Evaluating User Satisfaction with Music Discovery. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (Ann Arbor, MI, USA) (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 55–64. doi:10.1145/3209978.3210049
  - [41] Anja Nyland Hagen. 2022. Datafication, Literacy, and Democratization in the Music Industry. *Popular Music and Society* 45, 2 (2022), 184–201. doi:10.1080/03007766.2021.1989558
  - [42] David Hesmondhalgh. 2019. Have Digital Communication Technologies Democratized the Media Industries? In *Media and Society* (6 ed.). Bloomsbury Academic, London, 101–120. doi:10.5040/9781501340765.ch-006
  - [43] David Hesmondhalgh. 2022. Streaming's Effects on Music Culture: Old Anxieties and New Simplifications. *Cultural Sociology* 16, 1 (2022), 3–24. doi:10.1177/

- 17499755211019974
- [44] David Hesmondhalgh, Raquel Campos Valverde, D. Bondy Valdovinos Kaye, and Zhongwei Li. 2023. *The impact of algorithmically driven recommendation systems on music consumption and production – a literature review*. Technical Report. Centre for Data Ethics and Innovation (CDEI). <https://www.gov.uk/government/publications/research-into-the-impact-of-streaming-services-algorithms-on-music-consumption-the-impact-of-algorithmically-driven-recommendation-systems-on-music-consumption-and-production-a-literature-review> Accessed: 2026-01-21.
  - [45] Nyi Nyi Htun, Elisa Lecluse, and Katrien Verbert. 2021. Perception of Fairness in Group Music Recommender Systems. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) (IUI '21). Association for Computing Machinery, New York, NY, USA, 302–306. doi:10.1145/3397481.3450642
  - [46] Cherie Hu. 2025. *Music Tech Ownership Ouroboros, 2025 edition*. <https://www.waterandmusic.com/music-tech-ownership-ouroboros-2025/> Accessed: 2026-01-21.
  - [47] International Federation of the Phonographic Industry (IFPI). 2025. *Global Music Report 2025: State of the Industry*. [https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/GMR\\_2025\\_State\\_of\\_the\\_Industry\\_Final\\_83665b84be.pdf](https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/GMR_2025_State_of_the_Industry_Final_83665b84be.pdf) Accessed: 2026-01-21.
  - [48] Lucas D. Intra and Helen Nissenbaum. 2000. Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society* 16, 3 (2000), 169–185. doi:10.1080/01972240050133634
  - [49] Catherine Johnson, Matt Hills, and Laurie Dempsey. 2024. An audience studies' contribution to the discoverability and prominence debate: Seeking UK TV audiences' 'routes to content'. *Convergence* 30, 5 (2024), 1625–1645. doi:10.1177/13548565231222605
  - [50] Michelle S. Lam, Mitchell L. Gordon, Danaë Metaxa, Jeffrey T. Hancock, James A. Landay, and Michael S. Bernstein. 2022. End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 512 (Nov. 2022), 34 pages. doi:10.1145/3555625
  - [51] Jin Ha Lee and Rachel Price. 2015. Understanding Users of Commercial Music Services through Personas: Design Implications. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. International Society for Music Information Retrieval, Málaga, Spain, 476–482. doi:10.5281/zenodo.232222
  - [52] Jin Ha Lee and Rachel Price. 2016. User experience with commercial music services: An empirical exploration. *Journal of the Association for Information Science and Technology* 67, 4 (2016), 800–811. doi:10.1002/asi.23433
  - [53] Tuck W. Leong and Peter C. Wright. 2013. Revisiting social practices surrounding music. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 951–960. doi:10.1145/2470654.2466122
  - [54] Clayton Lewis. 1982. *Using the 'Thinking-Aloud' Method in Cognitive Interface Design*. Technical Report RC9265. IBM Thomas J. Watson Research Center. <https://dominoweb.draco.res.ibm.com/reports/RC9265.pdf> Accessed: 2026-01-21.
  - [55] Rena Li, Sara Kingsley, Chelsea Fan, Proteeti Sinha, Nora Wai, Jaimie Lee, Hong Shen, Motahare Eslami, and Jason Hong. 2023. Participation and Division of Labor in User-Driven Algorithm Audits: How Do Everyday Users Work together to Surface Algorithmic Harms?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 567, 19 pages. doi:10.1145/3544548.3582074
  - [56] Yu Liang and Martijn C. Willemsen. 2022. Exploring the longitudinal effects of nudging on users' music genre exploration behavior and listening preferences. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) (RecSys '22). Association for Computing Machinery, New York, NY, USA, 3–13. doi:10.1145/3523227.3546772
  - [57] Harriet Marsden. 2024. *Sabrina Carpenter and Spotify conspiracy theories*. <https://theweek.com/culture-life/music/sabrina-carpenter-spotify-conspiracy-theories-algorithm> Accessed: 2026-01-21.
  - [58] Eleonora Maria Mazzoli. 2020. Online content governance: Towards a framework for analysis for prominence and discoverability. *Journal of Digital Media & Policy* 11, Regulating digital platform power (2020), 301–319. doi:10.1386/jdmp\_00027\_1
  - [59] Eleonora Maria Mazzoli and Damian Tambini. 2020. *Prioritisation Uncovered: The Discoverability of Public Interest Content Online*. Technical Report. Council of Europe. <https://edoc.coe.int/en/internet/11681-prioritisation-uncovered-the-discoverability-of-public-interest-content-online.html> Accessed: 2026-01-21.
  - [60] Fenwick McKelvey. 2016. The new attention factory: Discoverability and Canadian cultural policy. In *Canadian Centre for Policy Alternatives Monitor*, 23(2), 32–34. Accessed: 2026-01-21.
  - [61] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) (CIKM '18). Association for Computing Machinery, New York, NY, USA, 2243–2251. doi:10.1145/3269206.3272027
  - [62] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human-Computer Interaction* 14, 4 (2021), 272–344. doi:10.1561/11000000083
  - [63] Judith Moeller and Natali Helberger. 2018. *Beyond the Filter Bubble: Concepts, Myths, Evidence and Issues for Future Debates*. Technical Report. University of Amsterdam, Amsterdam.
  - [64] Marta Moscati, Darius Afchar, Markus Schedl, and Bruno Sguerra. 2025. Familiarizing with Music: Discovery Patterns for Different Music Discovery Needs. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '25)*. Association for Computing Machinery, New York, NY, USA, 63–72. doi:10.1145/3699682.3728333
  - [65] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE* 9, 2 (02 2014), 1–23. doi:10.1371/journal.pone.0089642
  - [66] Laurens Naudts, Natali Helberger, Michael Veale, and Marijn Sax. 2025. A Right to Constructive Optimization: A Public Interest Approach to Recommender Systems in the Digital Services Act. *Journal of Consumer Policy* 48 (March 2025), 269–296. doi:10.1007/s10603-025-09586-1
  - [67] Thao Ngo, Johannes Kunkel, and Jürgen Ziegler. 2020. Exploring Mental Models for Transparent and Controllable Recommender Systems: A Qualitative Study. 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, 183–191. doi:10.1145/3340631.3394841
  - [68] NME. 2024. *Why is Sabrina Carpenter's 'Espresso' taking over Spotify algorithms?* <https://www.nme.com/features/music-features/espresso-sabrina-carpenter-spotify-streaming-3757741>
  - [69] Donald A. Norman. 2013. *The Design of Everyday Things*. The MIT Press, Cambridge, MA.
  - [70] Raphaël Nowak. 2016. *Consuming Music in the Digital Age: Technologies, Roles and Everyday Life* (1 ed.). Palgrave Macmillan, London. viii+167 pages. doi:10.1057/9781137492562
  - [71] Raphaël Nowak. 2016. When is a discovery? The affective dimensions of discovery in music consumption. *Popular Communication* 14, 3 (2016), 137–145. doi:10.1080/15405702.2016.1193182
  - [72] Anne Oeldorf-Hirsch and German Neubaum. 2025. What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *New Media & Society* 27, 2 (2025), 681–701. doi:10.1177/14614448231182662
  - [73] Ofcom. 2018. *EPG Prominence: A Report on the Discoverability of PSB and Local TV Services*. Technical Report. Ofcom. [https://www.ofcom.org.uk/\\_data/assets/pdf\\_file/0026/116288/report-psb-local-tv-discoverability.pdf](https://www.ofcom.org.uk/_data/assets/pdf_file/0026/116288/report-psb-local-tv-discoverability.pdf) Accessed: 2026-01-21.
  - [74] Yesid Ospitia-Medina, Sandra Baldassarri, Cecilia Sanz, and José Ramón Beltrán. 2023. Music Recommender Systems: A Review Centered on Biases. In *Advances in Speech and Music Technology: Computational Aspects and Applications*, Anupam Biswas, Emile Wennekes, Alicja Wiczorkowska, and Rabul Hussain Laskar (Eds.). Springer International Publishing, Cham, 71–90. doi:10.1007/978-3-031-18444-4\_4
  - [75] Will Page and Chris Dalla Riva. 2023. *'Glocalisation' of Music Streaming within and across Europe*. Technical Report EIQ Paper No. 182. London School of Economics and Political Science (LSE) European Institute. <https://www.lse.ac.uk/european-institute/Assets/Documents/LEQS-Discussion-Papers/EIQPaper182.pdf> Accessed: 2026-01-21.
  - [76] Gourab K. Patro, Lorenzo Porcaro, Laura Mitchell, Qiuyue Zhang, Meike Zehlike, and Nikhil Garg. 2022. Fair ranking: a critical review, challenges, and future directions. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 1929–1942. doi:10.1145/3531146.3533238
  - [77] Savvas Petridis, Nediya Daskalova, Sarah Mennicken, Samuel F Way, Paul Lamere, and Jennifer Thom. 2022. TastePaths: Enabling Deeper Exploration and Understanding of Personal Preferences in Recommender Systems. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) (IUI '22). Association for Computing Machinery, New York, NY, USA, 120–133. doi:10.1145/3490099.3511156
  - [78] Lorenzo Porcaro, Emilia Gómez, and Carlos Castillo. 2022. Perceptions of Diversity in Electronic Music: the Impact of Listener, Artist, and Track Characteristics. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 109 (April 2022), 26 pages. doi:10.1145/3512956
  - [79] Lorenzo Porcaro, Emilia Gómez, and Carlos Castillo. 2024. Assessing the Impact of Music Recommendation Diversity on Listeners: A Longitudinal Study. *ACM Trans. Recomm. Syst.* 2, 1, Article 3 (March 2024), 47 pages. doi:10.1145/3608487
  - [80] Robert Prey. 2020. Locating Power in Platformization: Music Streaming Playlists and Curatorial Power. *Social Media + Society* 6, 2 (2020), 2056305120933291. doi:10.1177/2056305120933291

- [81] Rebecca Salganik, Valdy Wiratama, Adelaida Afilipoaie, and Heritiana Ranaivon. 2024. Navigating Discoverability in the Digital Era: A Theoretical Framework. In *2nd Music Recommender Systems Workshop (MuRS 2024)*, CEUR Workshop Proceedings, Vol. 3787. CEUR-WS.org, Aachen, Germany, 1–10.
- [82] Michela Santangelo, Valentina Persici, Luca Caricati, Paola Corsano, Reyna L. Gordon, and Marinella Majorano. 2024. The adaptation and validation of the Goldsmiths Musical Sophistication Index (Gold-MSI) in Italian: The Gold-MSI-IT. *Psychology of Music* 52, 4 (2024), 385–401. doi:10.1177/03057356231204855
- [83] Markus Schedl, Peter Knees, Brian McFee, and Dmitry Bogdanov. 2022. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, 927–971. doi:10.1007/978-1-0716-2197-4\_24
- [84] Nick Seaver. 2019. Captivating algorithms: Recommender systems as traps. *Journal of Material Culture* 24, 4 (2019), 421–436. doi:10.1177/1359183518820366
- [85] Dougal Shakespeare, Victor Chareyron, and Camille Roth. 2025. Reframing the filter bubble through diverse scale effects in online music consumption. *Scientific Reports* 15, 1 (Feb. 2025), 4071. doi:10.1038/s41598-024-75967-0
- [86] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 433 (Oct. 2021), 29 pages. doi:10.1145/3479577
- [87] Ignacio Siles, Andrés Segura-Castillo, Mónica Sancho, and Ricardo Solís-Quesada. 2019. Genres as Social Affect: Cultivating Moods and Emotions through Playlists on Spotify. *Social Media + Society* 5, 2 (2019), 2056305119847514. doi:10.1177/2056305119847514
- [88] Ignacio Siles, Andrés Segura-Castillo, Ricardo Solís, and Mónica Sancho. 2020. Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society* 7, 1 (2020), 2053951720923377. doi:10.1177/2053951720923377
- [89] Spotify. 2025. *Loud & Clear*. <https://loudandclear.byspotify.com/> Accessed: 2026-01-21.
- [90] Spotify Italia. 2025. “Non succede, ma se succede... Ed è successo. Anche quest’anno la playlist di Sanremo 2025 su Spotify è la più ascoltata al mondo.” [Instagram post]. <https://www.instagram.com/spotifyitaly/p/DGBK8v3onbB/>
- [91] Spotify Newsroom. 2023. *Celebrating 10 Years of Spotify in Italy and Poland*. <https://newsroom.spotify.com/2023-03-01/celebrating-10-years-of-spotify-in-italy-and-poland/> Accessed: 2026-01-21.
- [92] Spotify Newsroom. 2024. *Culture Next 2024: The Major Gen Z Trends That Are Shaping Audio Streaming*. <https://newsroom.spotify.com/2024-11-04/culture-next-2024-the-major-gen-z-trends-that-are-shaping-audio-streaming/> Accessed: 2026-01-21.
- [93] Spotify Newsroom. 2025. *Spotify Takes Instagram Sharing to the Next Level with Audio Previews and Real-Time Listening Notes*. <https://newsroom.spotify.com/2025-08-21/spotify-takes-instagram-sharing-to-the-next-level-with-audio-previews-and-real-time-listening-notes/>
- [94] Statista. 2025. *Number of music streaming subscribers worldwide from the 1st half of 2019 to 4th quarter 2024*. <https://www.statista.com/statistics/669113/number-music-streaming-subscribers/> Accessed: 2026-01-21.
- [95] TikTok & Luminate. 2025. *Music Impact Report*. Technical Report. TikTok, commissioned with Luminate. <https://sf16-sg.tiktokcdn.com/obj/eden-sg/fuuvieh7uvzpsH/Music%20Impact%20Report%20-%20Final.pdf> Accessed: 2026-01-21.
- [96] TikTok Newsroom. 2023. *Add to Music App launches in partnership with Major Music Streaming Services*. <https://newsroom.tiktok.com/en-us/add-to-music-app-launches-in-partnership-with-major-music-streaming-services>
- [97] UNESCO. 2020. *France-Québec Mission on the Discoverability of Francophone Cultural Content Online*. <https://www.unesco.org/creativity/en/policy-monitoring-platform/france-quebec-mission-discoverability-francophone-cultural-content-online> Accessed: 2026-01-21.
- [98] Robin Ungruh, Karlijn Dinnissen, Anja Volk, Maria Soledad Pera, and Hanna Hauptmann. 2024. Putting Popularity Bias Mitigation to the Test: A User-Centric Evaluation in Music Recommenders. In *Proceedings of the 18th ACM Conference on Recommender Systems (Bari, Italy) (RecSys '24)*. Association for Computing Machinery, New York, NY, USA, 169–178. doi:10.1145/3640457.3688102
- [99] Angelina Wang, Vikram V Ramaswamy, and Olga Russakovsky. 2022. Towards Intersectionality in Machine Learning: Including More Identities, Handling Underrepresentation, and Performing Evaluation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (Seoul, Republic of Korea) (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 336–349. doi:10.1145/3531146.3533101
- [100] Carla Willig, Wendy Stainton Rogers, Gareth Terry, and Nikki Hayfield. 2017. Thematic Analysis. In *The SAGE Handbook of Qualitative Research in Psychology*. SAGE Publications Ltd, London, 17–36.
- [101] Brahim Zarouali, Sophie C. Boerman, and Claes H. de Vreese. 2021. Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics* 62 (2021), 101607. doi:10.1016/j.tele.2021.101607



## A A PRIMER ON THE CURRENT STATE OF THE ITALIAN MUSIC MARKET

Italy offers a prime example of a music market where strong domestic loyalty is increasingly shaped by global digital trends, such as short-form video platforms. Indeed, the Italian market remains unusual in the European context for its reliance on domestic production. According to a 2023 report on the “glocalisation” of music streaming within and across Europe by Page and Dalla Riva [75], the share of domestic artists in Italy’s top 10 songs rose from 30% in 2012 and 2017 to 70% in 2022, with the top six all by Italian artists. The 2025 report of the Federation of the Italian Music Industry (Federazione Industria Musicale Italiana – FIMI) [34] confirms this trend, showing an increase from 59% in 2014 to 84% in 2024.

Moreover, Spotify’s Loud and Clear report [89] notes that in 2025 Italian entered, for the first time, the eight most royalty-generating languages, surpassing the \$100 million mark following a 23% annual growth. In terms of discovery, TikTok is the leading platform in Italy, with nearly twice as many monthly active users as Spotify [75]. Luminate’s Global Music 360 survey for the first quarter of 2025 (reported in the Tuesday Takeaway Newsletter) further highlights the growing importance of short-form video, while also showing

that Italian Gen Z listeners are nearly 2.5 times more likely to discover music through Twitch and video games.

Genre preferences have also shifted. In 2019, the IFPI Music Listening Report [33] listed international and local pop as the most listened-to genres, followed by rock, singer-songwriters, dance/house, and hip hop/rap. This panorama changed drastically: as noted by Melanie Parejo (Spotify Head of Music for Southern and Eastern Europe), “Back in 2013, the most-streamed genre in Italy was rock, with Italian rap not even in the Top 10. But in 2022, Italian rap was the most-streamed genre in the country” [91]. Finally, Spotify’s explicit connection with the *Festival di Sanremo*, the Italian song festival par excellence, illustrates the interplay between platform and tradition. Spotify Italy was launched in 2013 during the festival, and in both 2024 and 2025, according to an Instagram post by Spotify Italy [90], the Sanremo Official playlist became the most-listened playlist worldwide.

## B CODING SCHEME

This appendix presents the coding scheme derived from the thematic analysis described in Section 4.5. Each table (Tables 3, 4, 5, 6) corresponds to one of the four themes, adopted deductively from the work of DeVos and colleagues [25].

**Table 3: Themes identified as “Exposures and Experiences”, capturing participants’ prior exposures to and experiences of recommender system behaviors, focusing on what users have observed, encountered, or felt as a direct result of interacting with music recommendation.**

Theme	Sub-Theme	Description	Count
<b>Relevant</b>		Users feel the recommendations align with their preferences, listening history, and evolving tastes.	36
	Recent / frequent	Recommendations that accurately reflect a user’s most recent or frequently played music.	20
	New releases	Recommendations of new music from artists or genres that match a user’s tastes.	7
	Similarity / familiarity	Recommendations that are similar to what a user already likes, even if not recently played.	9
<b>Irrelevant</b>		Users encounter recommendations that are surprising, random, or unhelpful.	53
	Minimal / accidental	Recommendations based on brief or unrepresentative listening interactions.	11
	Disliked / unfamiliar	Recommendations of music that users explicitly dislike or are completely unfamiliar with.	27
	Repetitive / stale	Recommendations that consistently feature the same artists, songs, or content.	10
	Outdated	Recommendations that align with the user’s past listening habits but are no longer relevant to their current tastes.	5
<b>Internal mechanism</b>		Users’ experiences with how the platforms generate recommendations and influence their music discovery.	54
	Serendipity / broadening	Recommendations introduce users to genuinely new or niche artists and songs they enjoy.	16
	Confinement / limited	Recommendations mainly propose known artists or genres, limiting new discoveries.	5
	Promotion / visibility	Recommendations seem to actively promote certain artists or content, possibly due to external factors, regardless of personal preferences.	33
<b>External influences</b>		Users’ observations of how outside factors like social media trends and media influence recommendations.	11
	Social media / virality	Music trends from social media platforms directly influence the recommendations users receive.	3
	Cultural media	Music from cultural media like TV series, movies, or video games influences what becomes prominent on the platform.	8
<b>UX / UI</b>		Users’ descriptions of their interactions with the app’s interface, including navigation and habits related to how music is presented.	37
	Navigation / routine	A user’s typical starting points and methods for navigating the app.	24
	Platform feature	Users’ engagement with and perception of specific app features designed for music discovery.	8
	Overall impression	General comments about the app’s visual layout, ease of use, and overall impression.	3

**Table 4: Themes identified for “Expectations and Values”, capturing participants’ anticipation of how a recommender would behave in a particular situation, forming expectations about recommender system behavior.**

Theme	Sub-Theme	Description	Count
<b>Temporal expectations</b>		Users’ expectations regarding the recency and frequency of the music recommended.	16
	Recent / frequent	Users expect recommendations to prioritize and reflect their most recent or frequently played music, acting as a continuation of their current listening habits.	4
	New releases	Users anticipate that recommendations will effectively highlight and present newly released music from artists or genres that match a user’s tastes.	12
<b>Preferences</b>		Users’ expectations about how recommendations cater to their individual tastes and the consistency of its recommendations.	19
	Similarity / familiarity	Users expect recommendations that align with their established musical preferences, offering content similar to what they already know and enjoy.	8
	Inconsistent / unexplained	Users encounter recommendations that feel out of place or don’t align with their perceived listening history.	11
<b>Scope</b>		Users’ perceptions of the range and breadth of music offered by the recommendations.	16
	Confinement / limited	Users feel recommendations keep them within a narrow range of genres or artists.	2
	Promotion / visibility	Users observe that recommendations seem to prioritize and promote mainstream artists or popular genres, sometimes at the expense of genuine personalization or niche discovery.	8
	Serendipity / broadening	Users desire recommendations to introduce them to genuinely new and diverse music, expanding their tastes beyond their typical listening habits in a surprising yet appealing way.	6
<b>Contextual</b>		Users perceive that external factors and real-world contexts influence the music they are recommended or choose to listen to.	6
<b>Users’ role</b>		Users’ active and passive behaviors in interacting with recommendations.	25
	Active / intentional	Users describe deliberate actions they take to find new music, such as searching or exploring specific platform features, often wanting the recommendations to better support these efforts.	9
	Passive / reliance	Users prefer effortless music discovery and rely on the recommendations to present relevant content without much effort on their part.	8
	Attraction / engagement	Users are drawn to specific elements that prompt them to interact with recommendations, such as appealing visual cues or a track’s immediate impact.	8

**Table 5: Themes identified for the “Folk Theories of Algorithms”, capturing participants’ mention of folk theories about why recommenders behave in the ways that they do, which informed their search and sensemaking process.**

Theme	Sub-Theme	Description	Count
<b>Commercial / promotional</b>		Users theorize that recommendations are influenced by commercial interests, payments, or promotional deals, leading to certain artists or content being pushed more prominently, often regardless of personal preference.	21
	Economic incentives	Users speculate that artists or labels might directly pay or have agreements with the platform for increased visibility.	6
	Platforms’ interests	Users believe the recommender’s behavior is tied to the platform’s need to generate revenue, leading it to promote widely popular artists to ensure engagement and broad appeal.	15
<b>Popularity / virality</b>		Users believe recommendations prioritize content that is currently popular, trending, or has gone viral, suggesting a reactive mechanism to widespread engagement rather than individualized taste.	17
	Social media	Users frequently cite social media, especially TikTok, as the primary source of viral content that influences visibility on music streaming platforms.	7
	High listenership	Users believe recommendations actively promote content that has already gained significant listenership, reinforcing existing popularity instead of focusing on novel discoveries.	10
<b>User Data / behavior</b>		Users believe recommenders learn directly from their past listening habits, search history, and interactions to provide personalized recommendations.	40
	Direct	Users believe recommenders primarily analyze their explicit listening history, like how often they listen to certain songs or artists.	24
	Inferred	Users believe recommenders can infer their broader musical taste, even without explicit listening, and can identify if they are a “niche” listener.	7
	Contextual / relational	Users believe recommenders also consider external contextual information, such as their friends’ listening, geographical location, or broader trends, to inform recommendations.	9

**Table 6: Themes identified as “Conceptions of Bias and Harms”, capturing participants’ definition and identification of recommender bias based on, e.g., the intentions behind the design, the potential for imposition, and the priorities of recommender systems.**

Theme	Description	Count
<b>Reinforcement of societal biases</b>	Users observe that recommendations seem to reflect and amplify existing societal biases related to gender, race, and niche genres.	25
<b>Commercial imperatives</b>	Users believe that recommendations are driven by financial motives, such as maximising engagement and promoting artists with major label backing.	30
<b>Confinement / limited discovery</b>	Users are concerned that recommendations confine their discovery by primarily recommending content similar to what they already consume, which limits their exposure to new or diverse artists and genres.	27