To be published in the proceedings of the 9th AAAI/ACM conference on Artificial Intelligence, Ethics, and Society (AIES) 2026.

AI-tribution: Enabling Training Data Attribution for Users of Generative Music Models

Julia Barnett[1*] **Nathan Pruyne**[2], **Nicholas Diakopoulos**[1], **Bryan Pardo**[1],

[1]Northwestern University
[2]Carnegie Mellon University

Abstract

Generative music models are growing in popularity, but users remain in the dark about the influences on their supposedly novel creations. Existing systems provide no mechanisms to identify highly similar songs or to help users learn how model outputs are influenced by training data. We present *AI*-tribution: a tool designed to inform users of generative music models about the training songs most similar to their generated outputs. We detail the interface, methodology, and a user study ($n = 26$) where musicians, media makers, and music consumers interact with the tool in 30-minute sessions. We analyze how the tool changes user knowledge of generated outputs, examine differences across user groups, and qualitatively analyze users' attitudes towards the tool. We find an overwhelmingly positive reception, show how the tool encourages deeper engagement with influences, and shed light on how different groups would use it in practice.

1 Introduction

Generative music models enable users to create "new" music at the click of a button. They are growing in popularity in terms of active user bases and models deployed (Copet et al. 2023; Garcia et al. 2023; Evans et al. 2025; Thickstun et al. 2024), and commercial models like Suno report tens of millions of users (UMG Recordings, Inc. v. Suno, Inc. 2024). Harms from these generative music models, e.g., copyright infringement and cultural appropriation (Barnett 2023), are keeping pace with this growth. The development of tools to empower users to mitigate these harms has not kept pace.

Every musician, from classical composers to sample artists to rock stars, draws inspiration from prior works. The statistical models of music learned by generative models are also influenced by prior works (Huang et al. 2018; Copet et al. 2023). A key difference between traditional creative inspiration and utilizing a generative music model is that a musician can cite their influences, while users of generative music models are left entirely in the dark. The black-box nature of the generation process precludes the option to engage with the influences upon the work. This research explores how user behavior and perception changes when presented with information about the influences upon their generated works and accompanying resources to learn about the music and artists similar to their generated output.

To evaluate this, we deploy a user study utilizing a generative model interface we built that presents users with the most similar works from the training data: *AI*-tribution, shown in the top row of Figure 1. The generative model, the automated method for identifying most-similar training data, and the interface are all fully functional, with details in Figure 1. We conduct a user study with three different user groups (musicians, media makers, and music consumers) to evaluate how user behavior changes when presented with information about training data influences, and present both quantitative and qualitative findings. It is important to study the effect of providing this information in three contexts: the system outputs memorized content from the training data, it outputs non-memorized content similar to training data, or it outputs content dissimilar to the training data. Generative music models do not generate copies or near-copies of training data with enough frequency to ensure every participant would be exposed to these situations (Barnett, Garcia, and Pardo 2024; Batlle-Roca et al. 2024). Therefore, despite the system being fully functional, we run this experiment using a Wizard of Oz (WOz) approach (Dow et al. 2005), replacing generated content as-needed with copies and near-copies to ensure study participants are exposed to these situations. Unlike attribution methods that estimate a training example's contribution via the model's internals (e.g., via unlearning (Choi et al. 2025)), we identify *similar* training examples through direct comparison of training data to model output. This lets one add our approach to any existing system without retraining or access to model weights. It also best captures market substitution: the songs most likely to be displaced in the market, due to their similarity to the generated output are the ones presented to the user. This method is also causally grounded when similarity is measured only on the data the model was trained on (as done in this work), since we know the compared songs were in the training set.

Our work makes four key contributions. First, (1) we show how **different user groups utilize different resources** to understand their generated outputs when presented with them alongside their generations. The second is to (2) examine how having access to training data information **helps different types of users learn** about their outputs. Our third is to (3) assess how having information about training data

[*]Now at Apple.

attribution **affects user confidence** in their understanding of the outputs and **comfort level** with performing various activities using their outputs. Finally, through thematic analysis of short interviews we (4) assess **why users prefer access to training data** for various motivations and uses. All of our contributions are enabled by the **introduction of *AI*-tribution**, which is the first end-user-facing tool for music that enables training data attribution at time of generation.

2 Background and Related Literature

Memorization in Generative Models

Generative models of text are known to memorize passages from their training data (Carlini et al. 2022, 2019, 2021; Peng, Wang, and Deng 2023; De Wynter et al. 2023). Image models have also been shown to memorize training data (Somepalli et al. 2023), as have text-to-audio models (Bralios et al. 2024). Model developers of symbolic music generation acknowledge this to be an issue; Thickstun et al. (2024) state, "...we are certain that generative models of music are similarly capable of plagiarism...Therefore, any outputs of this model should be presumed to risk infringing on copyrighted material."

Developers of some high profile generative music models have performed small memorization analyses in their research papers; prompting MusicGen with 10 seconds of audio resulted in 0.5-2% of generated outputs being memorized training data (Copet et al. 2023). Upwards of 10% of MusicLM's generations were approximate matches to training data (Agostinelli et al. 2023). Other works have enabled watermarking of training data in generative music models for attribution purposes (Epple et al. 2024; Yang et al. 2025). To our knowledge, methods to expose most-similar training data to the end user have not been pursued by the developers of many well-known music generation models (e.g., Google Magenta RealTime (Lyria Team et al. 2025), Suno, Udio). One possible reason is they are produced by for-profit companies, many of whom are currently engaged in lawsuits over copyright infringement of training data (UMG Recordings, Inc. v. Suno, Inc. 2024; UMG Recordings, Inc. v. Uncharted Labs, Inc 2024). Therefore, their interests do not align with developing such tools. While there have been attempts to mitigate potential copyright issues by use of open-source data (e.g., Stability Audio Open (Evans et al. 2025)), or by giving creators ways to opt-out of having their data used in training (e.g., Glaze, in the image domain (Shan et al. 2023)), they do not address the ignorance of the source material on which the generated music is based.

Harms from Uninformed Creation

Memorization is a capability and artifact of generative models, but not a harm in and of itself. Sampling, remixing, and interpolation are common practices in music making, however, they are used to explicitly give nods to or be in conversation with their source material. The harm occurs when users are not aware of memorization taking place, or when the work echos (in form, style, or content) existing work without directly copying it. We call this **uninformed creation**. When uninformed creation occurs, users may unknowingly appropriate songs or cultures (Agostinelli et al. 2023), or unknowingly infringe upon the copyright of the rights-holders of the song that has been memorized (Samuelson 2023). Other user-specific harms can be a result of not centering the user at the creative process (Barnett 2023): users can experience a loss of agency and authorship (Huang et al. 2020; Frid, Gomes, and Jin 2020), or stifling of creativity (Zhao et al. 2020; Huang et al. 2020; Chemla–Romeu-Santos and Esling 2022) while using these models.

Over time, this sidelining of the user may lead to broader cultural harms. Prior to generative models, borrowing a phrase or a sample from an earlier work meant listening to that music and learning the name of the artists that made it. For many artists, this is the first step in an exploratory process that leads to a deeper understanding of where their own creations fit into a larger cultural landscape. While models like Stability Audio Open avoid copyright issues, they also do not allow for this kind of learning and growth on the part of the artist using the tool. This may lead to an overall homogenization of music (Vanka et al. 2024) and a generation of creators who have been stripped of the creative process.

There is evidence from many other domains (Bergquist et al. 2023) that providing users with key information can positively nudge behavior. This has been found in domains as disparate as car driving (Lee-Kan et al. 2024) and meal selection (Chan, McMahon, and Brimblecombe 2021). We are, however, unaware of prior work that has explored the impact of providing information about most-similar works in the training data to the end users of generative music models. Recent works (Barnett, Garcia, and Pardo 2024; Batlle-Roca et al. 2024; Choi et al. 2025; Yin et al. 2022) developed and validated approaches to identify audio files in the training data of generative audio models that are highly similar to and replicated by the models, but none present most-similar training data to model users. In our work, we extend the approach introduced in our prior work (Barnett, Garcia, and Pardo 2024) by presenting this information to end users and studying its effect on their behavior and attitudes towards the use of the generated output.

3 Data and Methodology

Underlying Methodology

In our prior work we developed a methodology to systematically identify similar pieces of music audio in order to understand the influence of training data upon the outputs of generative music models (Barnett, Garcia, and Pardo 2024). We build on this approach due to the replicable nature of the methodology, its pairing with an open source model, and creators of that model's willingness to provide access to the training data. Though this methodology could be applied to the output of any generative music model, doing so previously required access to (a) outputs of models, (b) the training data of the model, and most importantly (c) the ability and initiative to implement their code. This reduced the potential user base down to generative model creators or highly motivated independent auditors. By implementing this system with an easily accessible model, we are able to provide

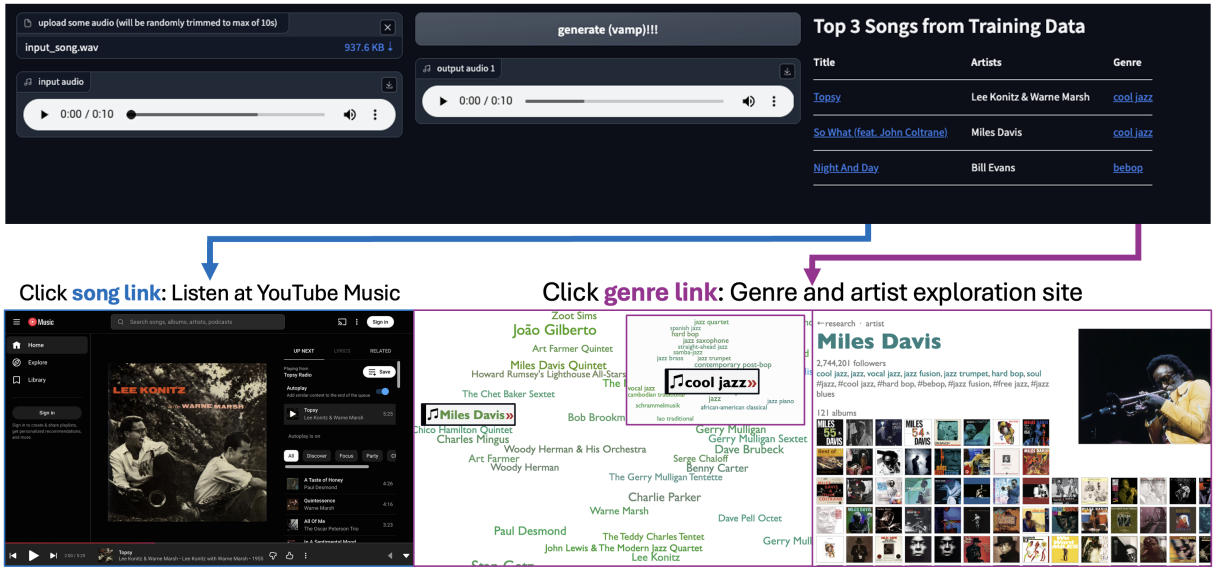

Figure 1: **Top row**: The interface of *AI*-tribution. Users can upload prompt audio from a song in the left panel and click "generate" in the middle panel to generate new music using a generative model (in this case, VampNet (Garcia et al. 2023)). Input files longer than 10 seconds will be trimmed to 10 seconds and all output files are 10 seconds long, an artifact of VampNet. When the output is finished, the third panel populates with the three most-similar songs from the model's training data. **Bottom left panel**: Each song name is hyperlinked with the YouTube Music link to that song at time stamp our model identified as being most similar. **Bottom middle and right panels:** The genres are provided as clickable links to the site Every Noise at Once (https://everynoise.com/), which contains every genre present in Spotify's internal database and the artists associated with that genre. The bottom middle panel shows related genres and artists in that genre for users to explore and listen; the bottom right shows an individual artist page on that site. We invite readers to watch a demo video at https://tinyurl.com/ai-tribution.

users with a ready-to-use system that requires no knowledge of model implementation or special access to model training data. We focus on similar songs, which as noted in the Introduction better capture market substitution risk.

Data and Models Used

Our interface, like the methodology it is built on, is model agnostic. However, to apply it to a *particular* model requires access to the model's (1) training data and (2) outputs. Once it is built for one model with a sufficiently large attached training dataset, **users can use outputs from any generative model and find the most similar songs in our data source**. Though using it in this way does not necessarily include songs other models may directly be memorizing or infringing upon, it still enables users to learn about artists and genres similar to their generations.

We built the interface on VampNet (Garcia et al. 2023), due to it being open source, the model's weights being publicly available, and the creators' willingness to share with us information about the training data. VampNet is a generative music-to-music model trained on 795k songs collected from the internet, which means it takes music audio as input and outputs (ideally) new music audio. We use both VampNet (the model) and a subset of VampNet's training data.

To identify similar songs from the training data, we use an embeddings-based similarity approach (Barnett, Garcia, and Pardo 2024), which extends the methodology from Somepalli et al. (2023) in the image domain which uses a "split-product" measure for similarity that breaks embedded feature vectors into chunks to compare localized features (as opposed to one vector representing the entire image). In our work, we embed each audio file into 3-second chunk vectors of Contrastive Language-Audio Pretraining (CLAP) embeddings (Wu et al. 2023), the use of which we validated using both quantitative metrics and a human listening study (Barnett, Garcia, and Pardo 2024). We upload all of the embeddings and musical metadata (e.g., song name, artist, genre) to a vector database so that we can efficiently identify similar matches. We then measure the cosine similarity between 3-second clips of the generated piece and the training data, and return the top three matches.

4 Experimental Design

Design Rationale

Because we want to understand how users behave with and without our tool, we use a within-subject study design and ensure each user experiences first the control condition: only the audio output of the model and then the test condition: audio output of the model accompanied by *AI*-tribution information. We were concerned that conducting a study where some people only had *control* and others only had *test* would confound results with differences between people, masking the effect of our tool, and require a substantially larger set of participants. Since participants generate only six songs, we did not expect fatigue to be a factor and did not counterbal-

ance ordering of control and test conditions, as participants might remember the outputs from the tool in the first phase if they did not receive them in the second phase.

User Groups and Recruitment

For this study we recruited 26 participants for 30 minute sessions, either over Zoom ($n = 21$) or in person ($n = 5$). We recruited across the United States from three self-reported user groups: (a) musicians, (b) non-musician media makers (e.g., podcasters, vloggers, filmmakers), and (c) music consumers. Our final user pool consisted of $n = 6$ musicians, $n = 10$ media makers, and $n = 10$ music consumers. All three groups were self identified based on the following definitions provided in a pre-screening survey regardless of professional status:

- Musician: someone who creates their own music.
- Non-musician media maker: someone who creates media that might use music such as podcasts, vlogs, or films.
- Music consumer: does not create music or media.

The musicians in our study had all composed music and took their status as a musician seriously (e.g., not hobbyists), though professional status varied. In terms of the Ollen Musical Sophistication Index (Zhang and Schubert 2019), all musicians identified as a "Serious amateur musician" ($n = 2$), "Semiprofessional musician" ($n = 1$), or "Professional musician" ($n = 3$). Importantly, no one identified as "Amateur musician" or lower items on the index. Genres also varied by and within musicians and included for example jazz, rock, and Latin folk. Media makers included social media content creators and managers ($n = 3$), documentary and filmmakers ($n = 2$), film editors ($n = 2$), podcasters ($n = 2$), and a Twitch streamer ($n = 1$). We ensured there was no overlap among the three categories: no media makers were also musicians and vice versa.

We recruited these participants primarily through various departments and class lists at the first author's home university, such as the music and film departments. We also recruited musicians in the city local to the first author, as well as media makers through a snowball recruitment method (e.g., film editors in LA and YouTube streamers). We paid each participant $20 in Amazon gift cards for their time, an estimated hourly wage of $40/hour, which is far greater than any minimum wage in the United States. Our study was in line with institutional review board standards and determined to be exempt by the first author's university IRB office; we had a thorough consent form and ensured participants knew they could quit at any time without penalty.

Data Preparation

We had three categories of outcome that interested us: those where the three most-similar songs returned by the system had *high* similarity, *moderate* similarity, or *low* similarity to the generated music. The generative model, the automated method for identifying most-similar training data, and the interface of *AI*-tribution are all fully functional; however, due to the stochastic nature of generative music models, they do not reliably produce copies or near-copies of training examples. If we were to use a model "in the wild" for this study, we would have no way to anticipate the expected behavior of the outputs and many participants would not have been exposed to all three outcomes. To avert this, we employed a Wizard of Oz (WOz) design (Dow et al. 2005). The input prompt audio was provided to participants, as was the pre-generated output audio that participants were going to "generate." Importantly, the participants did not know that they were not generating their songs on-the-fly at run time. We do not believe the WOz nature of this study made the experience less accurate, seeing as the experience itself was left largely untouched aside from the input and output songs chosen for the user; aside from that fact and the case for *low* songs detailed below, the user experience was largely the same as it would be in the wild. We detail the ways in which we manipulated the data in Table 1.

Wizard of Oz Design Data Manipulation		
Similarity Cat.	**Output Vamp**	**Output Top 3 Songs**
High	Manual Creation	No Manipulation
Moderate	No Manipulation	No Manipulation
Low	No Manipulation	Bottom of Top 10,000

Table 1: Description of data preparation for the WOz design. *High* similarity songs had manually created output vamps, but the output top three songs were returned as normal from our methodology. *Moderate* similarity songs were not manipulated. *Low* similarity songs had no manipulation on the vamps, but we instead chose songs from the bottom of the top 10,000 most similar rather than the top three to display.

The three most-similar training data songs were selected automatically by the system and validated by our author team. For *high* songs, one of the top three songs contained a substantially similar song (i.e., easily noticeable to a listener) at the linked time stamp. These were designed as examples of training data memorization in generative music; we created these manually to ensure the "copied" song was distinctively present. The goal of these *high* songs was to assess user behavior when the interface returns a song extremely similar to the generated music. The *moderate* songs had three top songs that sounded similar, but would likely not be mistaken for exact copies, e.g., a similar drum solo in a rock song. VampNet only outputs songs with extremely similar songs in the training data about 3% of the time (Barnett, Garcia, and Pardo 2024), however moderate similarity to training data commonly occurs in the wild. Finally, *low* songs had a set of "top three songs" that did not sound similar to the generated song (they were among the bottom of the top 10,000 most similar utilizing our method, the maximum limit our vector store returns); the point of including low-similarity songs was to assess blind faith in the tool and how users would perceive the information the tool provided if the music did not sound similar to human ears.

We prepared 15 sets of 10-second input song excerpts, 10-second generated output songs, and their identified top three songs from the training data for this experiment (for *low* similarity songs, these were instead songs from the bottom of the top 10,000 most similar). We also ensured the input song used to prompt the model was never an option

in the top three similar songs so as to not add a confounding variable. We had $n = 5$ input songs for each of the three similarity categories (*high*, *moderate*, and *low* similarity). All input songs were pulled from a subset of VampNet's training data. Half ($n = 8$) of the songs were from a subset of jazz and blues songs, and half ($n = 7$) were from a subset of pop and orchestral music. Each participant saw one random song from each category (*high*, *moderate*, *low*), and each participant saw at least one each of a jazz/blues song and a pop/orchestral song. This is detailed in Appendix Table 3.

All sessions were recorded for transcription and data analysis purposes. This allowed us to confirm the order in which participants generated songs, analyze participant behavior while using the tool, and transcribe the post-test interviews. We analyzed changes in the answers users gave regarding genres and artists similar to the outputs, changes in confidence levels, changes in how similar participants thought the generated songs were to existing songs, and changes to comfort level with different uses of the generated songs. We also conducted an inductive thematic analysis (Braun, Clarke, and Hayfield 2022; Terry et al. 2017) of responses during interviews. We started by open coding the interview transcripts (Khandkar 2009), then used axial coding (Vollstedt and Rezat 2019) to synthesize initial codes across interviews into coherent conceptual themes across user groups. We performed this iteratively, and support our findings and themes with quotes when appropriate. We analyze at the category level (*high, moderate,* and *low*). We quote individuals in the text using anonymous monikers (e.g., P7), but otherwise analyze at the group level (*musicians, media makers,* and *music consumers*) and compare across groups.

Session Protocol and Interview Design

Each participant was provided a set of three input songs to use in our interface and prompt the generative model. They did not know the generative model's behavior was predetermined, nor did they know which song would produce high or low similarity results from *AI*-tribution. We randomized the order in which the files appeared in participants' file folders, and participants were able to select their own order to input songs. Each order (e.g., [*high, moderate, low*], [*low, high, moderate*]) was seen roughly the same number of times (on average 4.3 times), counterbalancing this factor.

Participants had individual sessions that averaged 30 minutes in length. We started each session by making sure the participant had access to the files, interface, and answering any questions they had about the consent form (5 minutes on average). The bulk of the session consisted of the participant using our interface and answering a series of questions about their generated outputs (20 minutes on average). The questions asked in this survey defined the context of the tasks we were asking participants to put themselves in the mindset of; participants could go back and forth between the survey and the interface when answering the questions. The questions asked via survey covered genre identification, artist identification, confidence level for those answers, and comfort level with performing various use-cases (using in a YouTube video, using in a podcast, or releasing on Spotify). The survey was iteratively designed with feedback from five other independent HCI researchers to ensure the framing of the questions was responsive to our research questions. A full list of the survey questions can be found in the Appendix A. After the interface portion of the study, we conducted short (average five minutes) post-test interviews with participants.

Control: Pre-Tool Phase: At the beginning of the session we had participants download the audio files and upload them individually to the interface in an order they chose. They then generated a "new" (predetermined) output using the interface, but did not yet have access to the third panel displaying *AI*-tribution's songs from the training data. Instead, we provided them with access to YouTube Music and the genre exploration site Every Noise at Once. They then answered survey questions about each generated song.

Test: Post-Tool Phase: After the participants finished their third generated song, we moved onto the second phase of the study, where our tool presents users with information about similar songs from the training data to the user's generated output. To get participants comfortable with the tool, we first demonstrated how the tool worked using a generated output matched to an extremely similar song, so they could see the potential utility of our interface. We then explained that not every generated song would have something extremely similar in the training data and showed how the tool behaves when there is not a close match in the training data. Every participant saw the same examples in this demonstration.

We then gave participants access to the interface that included *AI*-tribution, and asked them to again generate three output songs, using the same inputs as before. This resulted in the same set of generated audio due to the WOz nature of the study. When participants asked, we told them the tool saved their outputs from the first phase (pre-tool phase) so they would have the same outputs in the second phase. For each generated output, participants saw a list of the the top three "most similar" songs. Song names were hyperlinked with the YouTube Music link to that song at the time stamp our model identified as being most similar. The genre of the song was provided as a clickable link to the site Every Noise at Once. This let users listen to the song and explore the genre attached to that song.

Post-test Interview: We ended each session with a short interview (on average 5 minutes) where we asked participants questions about their experience using the tool, motivation for answers on the survey, and other questions relevant to each group's likely use of the generated music. See the Appendix B for specific interview questions.

5 Results

In this section we will first (1) examine the degree to which users utilized the provided resources to assess how much more likely users are to evaluate their outputs with access to our tool. Next, we will (2) analyze the type of answers users gave pre- and post-using *AI*-tribution for artist identification to understand how our tool changed their understanding of the outputs (analysis on genre identification is in Appendix C). Finally we will (3) analyze their confidence and comfort in using the generated works in various tasks both pre- and post-using *AI*-tribution to gauge users' perceptions of their own knowledge gain from using our tool.

Frequency of use of information resources with/without *AI*-tribution										
	Genre Exploration		YouTube			Both Tools			At Least 1 Tool		
User Group	Without	With	Without	With	*p*	Without	With	*p*	Without	With	
All Participants	38%	38%	12%	**92%**	***	7%	**35%**	*	42%	**96%**	***
Musicians	33%	**50%**	17%	**100%**	**	17%	**50%**		33%	**100%**	*
Media Makers	20%	**40%**	10%	**90%**	***	0%	**40%**	*	30%	**90%**	**
Music Consumers	**60%**	30%	10%	**90%**	**	10%	**20%**		60%	**100%**	*

Table 2: Usage of the genre exploration tool Every Noise at Once and YouTube Music. Values display the percent of trials where participants used that tool at least once, with "Both Tools" displaying the % of trials where participants utilized both tools at least once. Bold numbers indicate the phase (without/with *AI*-tribution) in which more participants utilized the resource(s). Significance results displayed for paired sample t-tests, with $* < 0.05$, $** < 0.01$, and $*** < 0.001$.

Overall Usage of Resources

At the start of each session, participants were told they could access YouTube Music and Every Noise at Once. We also opened these sites in tabs on the computer, if the participant was in person, or sent the links to the sites in chat if their session was over Zoom. Despite this priming, most participants chose not to use these resources during the control phase of the user study when they were asked about what artist or genre the generated music most resembled (results shown in Table 2). Media makers had a strong tendency to speed through the first phase of this study without looking at any resources. Musicians cited an already-extensive knowledge base of genres, leading them to choose not to utilize these resources in the pre-tool phase. When participants encountered friction (e.g., not knowing where to start or what to search for) they did not use resources presented to them. However, **after being presented with access to the songs and links provided by *AI*-tribution** they were **more likely to leverage at least one provided resource** ($42\% \rightarrow 96\%$).

Music consumers, on the other hand, were more likely to access the genre exploration site to aid them in the first phase of this study: 60% used this at least once before gaining access to *AI*-tribution. Almost every participant accessed the YouTube Music links when they had access to *AI*-tribution, even though few used this site prior to seeing the *AI*-tribution information ($12\% \rightarrow 92\%$).

Participants rarely used both the genre exploration site and YouTube Music in the pre-*AI*-tribution phase, but it was more likely for participants to use both tools at least once with access to *AI*-tribution ($7\% \rightarrow 35\%$). Music consumers were more likely than the average participant to just use one or the other in both phases, and media makers were less likely to use either resource in both phases.

One participant (P9) described the difference between using the tools prior to having *AI*-tribution information and after: before, "I needed to be the one to initiate that process of looking for connections. Whereas the tool surfaced connections and put me as an evaluator...when I'm searching myself there was a thing like 'okay what if I just haven't searched for the right term?"'. Another music consumer (P24) noted how being guided in the genres to start with in *AI*-tribution helped them understand the piece: "I think the knowledge changed pretty significantly. I think I was going more off of vibes in the first portion, but then once you showed me stuff it gave me a starting point at least for where to explore and then map the genres and artists."

Changes in Genre and Artist Identification

We next want to examine how having access to information about training data attribution helps users learn about their generated outputs. We report artist results in the main text, and genre results are in the Appendix C. The first questions on the survey asked users to identify the genre and artist most similar to the generated output of the model. We categorized their answers for both control and test phase into categories. For both phases this includes:

- *Within top 3*: their answer was in the top three songs the tool will display (though they do not know this)
- *Not in top 3*: their answer was not in the top three songs the tool will display (though they do not know this)
- *I don't know*: user described some variation of not knowing the answer; typically wrote "I don't know"

For participants' answers after seeing *AI*-tribution, we retained the same categories (labeling the answer as "no change" if it stayed in the same category) with one addition:

- *More descriptive*: added an adjective to their description or additional name (e.g., rock $\rightarrow$ punk rock; Miles Davis$\rightarrow$Miles Davis, Chet Baker)

We first analyze the changes in users' ability to identify artists whose work was highly similar to the generated music before and after having access to *AI*-tribution. The full output of the categorized answers users gave can be seen in Figure 2. This flow figure displays the first answer users gave and how (or if) they changed their answer based on information provided via *AI*-tribution. A breakdown of this figure for the three user groups can be found in Appendix Fig. 5.

High (similarity) song behavior was relatively consistent across user groups: when a generated song was extremely similar to one in the training data, participants changed their answer from either "I don't know" (50%) or something not in the top 3 (46%) to something in the top three displayed in *AI*-tribution (92%), typically the artist of the song that was extremely similar. In the instances where this did not occur

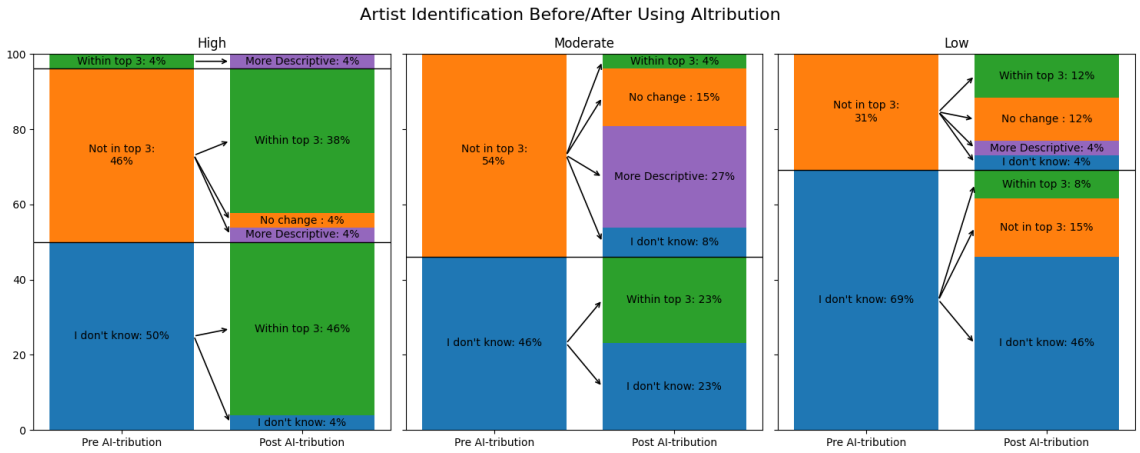

Figure 2: Variation of a Sankey diagram (Riehmann, Hanfler, and Froehlich 2005) displaying the distribution of answers given for artist identification and how they changed pre- (left bar) and post-access (right bar) to *AI*-tribution's top 3 songs in the training data. Split across low, moderate, and high behavior songs (in columns) and the three user groups (in rows).

($n = 2/26$), the participants did not use the YouTube Music links for the songs but instead just the genre links.

Low (similarity) song behavior revealed whether there was blind trust in the tool among our participants. For the low song, everyone either said "I don't know" (69%) or something not in the top 3 of the training data (31%) in the first pre-*AI*-tribution phase, and only 20% of users changed their answer to something displayed in the top 3 after viewing the outputs from *AI*-tribution. The vast majority of these were media makers; no musicians blindly trusted the tool, and most music consumers ($9/10$) maintained that they did not know or added an answer not on the tool. **Media makers were much more likely to put blind trust in the tool** and list its output for similar artists, either without listening to songs or deciding the tool knew better than they did.

Moderate (similarity) song behavior had the most diverse answers among the three. Though everyone still either named an artist not provided by *AI*-tribution (54%) or "I don't know," (46%) before being shown *AI*-tribution output, there were much more diverse answers once similar songs were suggested by *AI*-tribution. For those who had named an artist in the pre-phase that ended up not being among the top 3 displayed by *AI*-tribution, 50% became more descriptive in their answer (sometimes naming an artist on the tool in a list of others), 28% did not change in originally answering something off the tool, among other less frequent behaviors. Musicians maintained the most skepticism in keeping their "I don't know" answers, whereas both media makers and music consumers were similar in their diverse changes.

Musicians cited obscurity as a driving factor of change in answers: P1 noted, "There are a lot of obscure artists, and obscure subgenres. And I am familiar with a lot of that stuff but there's so much music out there. And there's so much stuff that I'm assuming this model has been trained on that I'm not aware of. Having those references...after hearing that clip it became pretty obvious to me." Media makers echoed this sentiment in their own words: "I don't have a bank in my head of all songs ever released. So it feels like what do I know about if this is similar to something else versus going 'okay these are the songs it's most similar to.'...there's a greater level of trust and it improves my literacy" (P7).

Changes in Confidence and Comfort Level

We next want to understand how having access to training data information affects user confidence in their understanding of their generated outputs, as well as their comfort level with using those outputs for various tasks (e.g., using it in a YouTube video or releasing it on Spotify). We analyze the changes in confidence scores for genre and artist, identifiability of the song to existing artists and songs, and comfort using the generated song for various purposes. All questions were asked on some variation of a 5-point Likert scale, with a maximum of 5 (e.g., extremely confident) and minimum of 1 (e.g., not at all confident). The changes in all scores can thus be calculated on the same 5-point scale, though answers to different types of questions are not necessarily 1:1 comparable. The aggregated results are displayed in Fig. 3, and results by user group can be seen in Appendix Fig. 7. The scores within panels split by a gray line are 1:1 comparable.

Across the board for all user groups, changes in answers were most dramatic for generated songs with high similarity to something in the training data. After having access to *AI*-tribution, users were (1) more confident in their genre and artist identification, (2) thought people would be more likely to recognize the artist if they heard the generated song and thought it was more similar to existing songs, and (3) were much less comfortable releasing the song on Spotify, using it in a YouTube video, or using it in a podcast.

Media makers had more extreme differentials for every question than both other groups—the tools had the most dramatic impact on their confidence and comfort levels across use-cases. They were also the only group to consistently become *more* comfortable with releasing the output on Spotify, using it in a YouTube video, and using in a podcast when the song had low-similarity. This may be because the training data not being similar confirmed for them they would

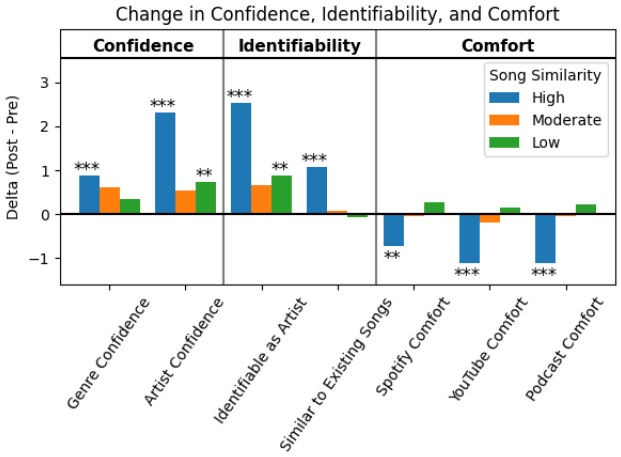

Figure 3: Changes in confidence scores, identifiability of artists and songs, and comfort using generated songs in various use-cases, calculated as post- minus pre-*AI*-tribution access. Bar colors from left to right indicate the change for high-, moderate-, and low-similarity songs. Positive bars are increases, negatives are decreases. Asterisks above/below the bars indicate significance scores for paired sample t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

not be infringing on someone's copyright. As P7 noted, "my biggest two things are (1) does it fit the vibe of what I'm trying to create, and (2) will I get copyright flagged for it." They elaborated that *AI*-tribution changed their experience of AI music generation: "It transformed the experience actually. I left feeling like I have a lot more confidence in if it's similar or not, and even when it was similar I felt grateful that I knew that and I didn't have false information. And when it was dissimilar, I felt very confident in the ways in which it was dissimilar, like I could trust my own judgment."

Musicians were less extreme. Their genre confidence for all three songs had almost zero change, meaning the tool did not dramatically change their evaluations. Their artist confidence changed more than genre, due to the specificity of the tool in identifying artists. They were the only user group to consistently become less comfortable with *both* high and moderate songs for all three activities (Spotify, etc.).

Music consumers had drastic changes for the high similarity songs, but relatively modest changes for the moderate and low songs. This is in line with this group consistently saying "I don't know" for answers outside of high similarity; they feel they have learned something, but they are still not as confident about it. Some music consumers described the reason for their general discomfort regardless of similar songs: "I think it's because someone can directly stand to profit off of them...Being able to see the attribution makes me realize how similar music can be. And even if it wasn't very similar to the attributed data, it was similar enough to the [input] often times that you were using to generate it, and that just made me feel really uncomfortable."

For ***high similarity songs***, users in all three groups noted how drastically their attitudes changed. One media maker (P2) described their changes in comfort level of releasing it in a YouTube video: "the model recognized the source pretty well. So that drastically changed my attitude to that. Went from very comfortable to very uncomfortable." Another media maker (P8) experienced a similar change: "initially I was like, 'this is not recognizable, maybe you could use it in a vlog or some form of a YouTube video,' but then after realizing how easy it is to find the real owner I was like actually no lets not even try that." A musician (P5) echoed this: "When I heard how identical they were...I can confidently say it's this artist and everyone will recognize it."

For ***moderate similarity songs***, there were mixed results in how comfort levels changed for using the generated songs. Comfort levels often remained unchanged after exposure to *AI*-tribution, but some participants noted how it actually increased their comfort level. One media maker (P2) took the lack of extreme similarity to increase their comfort with their creation: "The attribution model didn't find a very similar source, so in that case it let me know some of the inspirations...so that still seems normal to me and acceptable to me." A music consumer (P18) explained their thought process: "It's first how similar does the music actually sound to music that you can attribute to an artist. And then the other layer was how similar does that genre already feel to other songs in the genre. For example with the [moderate song], that genre felt very similar to me throughout artists, so it feels slightly more comfortable to use AI-generated music."

For ***low similarity songs***, media makers consistently became more comfortable using those outputs for various tasks. The core of this seemed to be confidence they would not get sued for using those songs. P7 described why: "When they don't sound anything alike I feel like I have more confidence in 'okay this song won't get copyright flagged' because nobody is going to mistake this for Frank Sinatra, that's just not gonna happen. So that leaves me feeling a little more confident in the tool rather than just not knowing where it's being pulled from." Another media maker (P10) more directly states, "Now I'm extremely comfortable using them because it doesn't sound like those songs at all. Not on an ethical basis, but on the basis of whether I'd be sued."

6 Thematic Analysis: Qualitative Findings

Every participant's session ended with an interview (see Appendix B) to discuss their thoughts and attitudes towards the tool. We now discuss trends observed across interviews.

Unanimous Preference for the Tool, but Varied Motivation and Function

Every participant expressed a positive reaction to the tool, and all said if they were to use a generative music model they would prefer access to a tool that exposed the training data. Media makers wanted it in their workflows, music consumers wanted it for entertainment and educational purposes, and musicians either wanted it available for non-musicians or to assist them in learning about new artists. Across all groups, the desire for attribution information was motivated by a need to regain control over uncertain or ethically ambiguous creative processes, though the function it served varied; protection from litigation for media makers,

sense-making for music consumers, and for musicians an avenue to bring non-musicians into traditional workflows.

Media makers noted they would always prefer knowing the sources behind songs, especially to avoid copyright infringement or accusations of theft. One media maker (P8) described the necessity of understanding sources: "as a video editor, if you're using any type of music, any type of asset, that you did not make...you want to know where it comes from." A film editor (P12) echoed this sentiment: "If you're trying to find something [to use], the more tools of access you have to locating where the source is coming from can help educate you to make a better decision about what you want musically." Media makers cited their own lack of musical literacy in desiring this tool: P14 describes, "I would want to use something that has the training tool attached to it rather than rely on my own musical intuition which is not very much." Their desire for information about similar sources was often rooted in a desire for risk mitigation surrounding the legal legitimacy of using this content.

Music consumers described having access to *AI*-tribution's training data information as more useful, fun, and important in light of their distrust of AI generated music. P16 described a desire to know the source of their generations: "I like the training data. I like to know what I'm getting, where it's coming from." P15 enjoyed use of the tool more than the generative model itself: "I personally liked having the data because I thought honestly that was the most fun part of it for me," a sentiment echoed by P18.

Their desire for training data attribution reflects an overall distrust of generative AI and a need to regain agency. Some music consumers described how this tool eased their concerns with AI generated music: "Using the artificial music makes me really uncomfortable, but the attribution tool is really helpful so I can be like 'Oh this is similar to that, and this is why,' it's much more of an educational tool" (P9). It also helped others realize their own concerns: "One thing that freaked me out was the last song—where I was like oh my god this is straight up taken from that Beach House song—without those inputs and checking them or without the training data attribution I don't think I would have known that" (P24). These music consumers were simultaneously entertained by attribution information and its learning opportunities, but sometimes unsettled by what it revealed.

Musicians noted that while they may not use it in their own workflows due to an overall resistance to AI generated music, they still wanted other people who were using those models to have access to *AI*-tribution. One musician (P22) succinctly noted "I think knowing the training data would be awesome." Another (P25) stated that even though they may not agree with the outputs, they can understand what non-musicians might believe: "It does help to see some of those top three songs from the data because it sort of tells you what other people might think." Others (P5) noted it transformed the AI music generation process to be more similar to musicians': "I like the idea of having the inspirations next to it. After all, every artist is inspired by someone else. In the guitar world everyone is inspired by Jimi Hendrix and we're not all paying royalties to Jimi Hendrix. But I think having the recognition is very important and very helpful."

Some of their views reflected a conditional endorsement: they do not want the generative music tools to exist at all, but if they do, they want the attribution information to be attached. The use of attribution information in this way enabled the users of generative models to enter in a long standing tradition of understanding musical influences.

Emergent Uses of Tool

This study elucidated unenvisioned use cases. One media maker (P7) described the tool as educational: "I felt like I got greater insight into the kinds of sounds I gravitate to more," detailing that they did not previously "have the verbiage to describe what [they] liked about it." Another media maker (P21) said: "I am someone who is not at all aware of the music genres properly. So, it's very useful. It's giving me the artist name, it's exposing me to new artists, it's exposing me to different genres." Participants used the tool both to learn about influences of their outputs case by case, and to build general musical literacy. This was not simply a function of the standalone resources; Table 2 shows participants rarely sought these out unprompted. The educational value came from the full system directing users to comparable songs.

Participants in all groups saw it as a **way to identify human created music they would rather use than what they generated**. One music consumer (P11) noted, "The first song from the training data was so similar to the generated music, I think I would use it, but attribute it to the training data rather than attribute it to the generated piece of music." One film editor (P13) noted they could identify a sound they liked via generative models, then find existing similar non-AI sounds: "It's a creative thing, and it's hard to nail down specifically what you're looking for. So the more things you have that are pointing you in the right direction, the more helpful it is." One musician (P26) echoed this thought for drawing composition inspiration: "If I have [*AI*-tribution], it'll be great because...it will help me in my investigation of finding bands." Across all user groups the attribution information was both a pathway to understanding more about their outputs and an avenue redirecting the user back to human-made works.

Music consumers and media makers both ascribe *AI*-tribution with providing them a "new vocabulary" to describe songs. P7 remarks, "I loved [*AI*-tribution] because I don't have a lot of music literacy, so to be able to get a peek behind the curtain...just to know more of the niche terms; I never would have come up with 'dutch pop' on my own if you left me to my own devices." One music consumer (P24) even said that having access to better genre information helped them identify similar artists: "I am bad at putting exact words to a song genre. So I think it was helpful to give me words but also to connect artists that I had a hunch would be in this category."

7 Discussion

GenAI Users Desire Attribution Information

Though reasons varied by user group, each of our participants desired training data attribution information. Media makers frequently commented on how useful this tool would

be in their workflows (P12, P13, P21). They were glad they had this tool to help them understand the music they may incorporate into their other creations, as well as avoid lawsuits or being flagged for copyright violations (P2, P7, P10). Music consumers tended to prefer using our tool for the fun of it—multiple participants said they found it more interesting than the generative music model itself (P15, P18, P24), and certainly more ethical (P9, P20). Musicians indicated they would much prefer to use a tool with this training data attribution over those without. Designers of generative audio tools (and perhaps other generative tools) must consider ways to **integrate attribution information into the interface or outputs of generative tools**.

The bulk of creativity support tools within generative AI are geared towards providing a functional or efficient improvement (Zhang et al. 2023; Louie et al. 2020; Jin et al. 2024). Few focus on showing users the broader cultural context to which their creations belong. Users want to understand if they are creating music that sounds similar to existing genres or artists, who those artists are, and in some cases they even want to use generative AI as a stepping stone to identify human artists whose work they want to use and properly credit. In an age where generative AI outputs are a viable market substitution for existing artist work, this tool puts those original artists who contributed to the training data back in conversation with the user who may have unknowingly created something similar to them. This principle extends beyond music generation; this work demonstrates a method for conveying information to end users that is actionable for providing cultural context of the outputs.

Present Attribution in Context, Not Isolation

One goal of this work is to encourage users to learn more about the cultural context of generated outputs. Participants frequently read aloud the names of genres and artists and remarked "I have no idea what that is." (e.g., P10, P18) and then proceeded to click the links and listen with their own ears to the most similar songs and genres to assess whether they agreed. Without access to resources where they could learn and assess for themselves, users would merely have one additional black box layer of outputs: a list of similar data points without knowing why or how they were flagged. Trust in the attribution functionality is built when they can decide for themselves if they agree with the identified similar music. Even media makers, who demonstrated the highest level of blind trust in the tool, were still more likely to engage with the information resources when they were presented alongside the outputs (see Table 2).

Effects of Uninformed Creation

Memorization of training data is a known phenomenon in generative models of all mediums, and generative music models are no exception. Paired with uninformed creation, memorization of training data can result in harms such as copyright infringement and cultural appropriation (Barnett 2023). We address both of these harms with this work. By exposing the information about the most similar songs in the training data, we enable users to check for themselves if the most similar songs would potentially infringe on someone's copyright, especially if they have never heard the song or artist before. If the works are not so substantially similar as to warrant copyright concerns, users can still learn about the genre their music falls into. Lalonde (2021) cites three potential harms surrounding cultural appropriation: nonrecognition, misrecognition, and exploitation; by providing users with resources about the genre and type of music they have just created, we help them place their creation into a cultural context, avoiding exploitation of a culture to which they may not belong. Without knowing anything about the most similar works, it is not possible to even begin to have the conversation about appropriation; we are giving users the opportunity to open that door to that conversation.

Limitations

AI-tribution was implemented using VampNet's training dataset (Garcia et al. 2023) of 795k songs. If a user were to upload generated music (e.g., from a different generative model) that is a memorized copy of a song outside the dataset, then *AI*-tribution could not return the most-similar song and the user would be under the false impression that they are not infringing.

A limitation of our user study is that we have a biased sample of participants who self-selected due to interest in participating in a research study. Users with different biases may ignore the results displayed by *AI*-tribution at higher rates than participants engaging in a user study under the watchful eye of a researcher. We have no data on usage of this tool "in-the-wild," and thus cannot comment on its efficacy outside of this research study. Additionally, the music generation was divorced from a particular task (e.g., scoring an action chase scene in a film) and instead existed as an out-of-context task (create a piece of music); a longitudinal study examining how users engage with *AI*-tribution in their own workflows would address this.

8 Conclusion

In this work we asked the question: How does user behavior change when presented with the most-similar training data, alongside outputs from a generative music model? To explore this, we built an interface, *AI*-tribution, that identifies the three most-similar songs, alongside the output from the generative model, along with hyperlinked resources to genre and artists for the most-similar songs. This information helped users identify when generated music was highly similar to music in the training data, learn about similar artists and genres, and become more confident in their own understanding of the generated music. The overwhelmingly positive reception to this inclusion of training data attribution information has important design implications for creativity support tools for generative AI systems. Future systems should expose this information to users to sate the desire users have of understanding the broader cultural context of their creations. *AI*-tribution is a tool that can mitigate the harms from uninformed creation, and help people become more informed and more conscientious users of generative music models. By presenting attribution information alongside generated outputs, we transform generative models into tools that help users learn about the cultural context that contributed to training the systems in the first place.

References

Agostinelli, A.; Denk, T. I.; Borsos, Z.; Engel, J.; Verzetti, M.; Caillon, A.; Huang, Q.; Jansen, A.; Roberts, A.; Tagliasacchi, M.; et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*.

Barnett, J. 2023. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 146–161.

Barnett, J.; Garcia, H. F.; and Pardo, B. 2024. Exploring musical roots: Applying audio embeddings to empower influence attribution for a generative music model. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*.

Batlle-Roca, R.; Liao, W.-H.; Serra, X.; Mitsufuji, Y.; and Gómez, E. 2024. Towards assessing data replication in music generation with music similarity metrics on raw audio. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*.

Bergquist, M.; Thiel, M.; Goldberg, M. H.; and Van Der Linden, S. 2023. Field interventions for climate change mitigation behaviors: A second-order meta-analysis. *Proceedings of the National Academy of Sciences*, 120(13): e2214851120.

Bralios, D.; Wichern, G.; Germain, F. G.; Pan, Z.; Khurana, S.; Hori, C.; and Le Roux, J. 2024. Generation or replication: Auscultating audio latent diffusion models. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1156–1160. IEEE.

Braun, V.; Clarke, V.; and Hayfield, N. 2022. 'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative research in psychology*, 19(2): 424–445.

Carlini, N.; Ippolito, D.; Jagielski, M.; Lee, K.; Tramer, F.; and Zhang, C. 2022. Quantifying memorization across neural language models. In *The Eleventh International Conference on Learning Representations*.

Carlini, N.; Liu, C.; Erlingsson, Ú.; Kos, J.; and Song, D. 2019. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX security symposium (USENIX security 19)*, 267–284.

Carlini, N.; Tramer, F.; Wallace, E.; Jagielski, M.; Herbert-Voss, A.; Lee, K.; Roberts, A.; Brown, T.; Song, D.; Erlingsson, U.; et al. 2021. Extracting training data from large language models. In *30th USENIX security symposium (USENIX Security 21)*, 2633–2650.

Chan, J.; McMahon, E.; and Brimblecombe, J. 2021. Point-of-sale nutrition information interventions in food retail stores to promote healthier food purchase and intake: A systematic review. *Obesity Reviews*, 22(10): e13311.

Chemla–Romeu-Santos, A.; and Esling, P. 2022. Challenges in creative generative models for music: a divergence maximization perspective. In *3rd Conference on AI Music Creativity (AIMC 2022)*.

Choi, W.; Koo, J.; Cheuk, K. W.; Serrà, J.; Martínez-Ramírez, M. A.; Ikemiya, Y.; Murata, N.; Takida, Y.; Liao, W.-H.; and Mitsufuji, Y. 2025. Large-Scale Training Data Attribution for Music Generative Models via Unlearning. *arXiv preprint arXiv:2506.18312*.

Copet, J.; Kreuk, F.; Gat, I.; Remez, T.; Kant, D.; Synnaeve, G.; Adi, Y.; and Défossez, A. 2023. Simple and controllable music generation. *Advances in Neural Information Processing Systems*, 36: 47704–47720.

De Wynter, A.; Wang, X.; Sokolov, A.; Gu, Q.; and Chen, S.-Q. 2023. An evaluation on large language model outputs: Discourse and memorization. *Natural Language Processing Journal*, 4: 100024.

Dow, S.; MacIntyre, B.; Lee, J.; Oezbek, C.; Bolter, J. D.; and Gandy, M. 2005. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing*, 4(4): 18–26.

Epple, P.; Shilov, I.; Stevanoski, B.; and de Montjoye, Y.-A. 2024. Watermarking Training Data of Music Generation Models. *arXiv preprint arXiv:2412.08549*.

Evans, Z.; Parker, J. D.; Carr, C.; Zukowski, Z.; Taylor, J.; and Pons, J. 2025. Stable audio open. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Frid, E.; Gomes, C.; and Jin, Z. 2020. Music creation by example. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13.

Garcia, H. F.; Seetharaman, P.; Kumar, R.; and Pardo, B. 2023. VampNet: Music generation via masked acoustic token modeling. In *Proceedings of the International Society of the Society of Music Information Retrieval (ISMIR)*.

Huang, C.-Z. A.; Koops, H. V.; Newton-Rex, E.; Dinculescu, M.; and Cai, C. J. 2020. AI song contest: Human-AI co-creation in songwriting. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.

Huang, C.-Z. A.; Vaswani, A.; Uszkoreit, J.; Shazeer, N.; Simon, I.; Hawthorne, C.; Dai, A. M.; Hoffman, M. D.; Dinculescu, M.; and Eck, D. 2018. Music transformer. In *Proceedings of The International Conference on Learning Representations*.

Jin, Y.; Cai, W.; Chen, L.; Zhang, Y.; Doherty, G.; and Jiang, T. 2024. Exploring the design of generative AI in supporting music-based reminiscence for older adults. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–17.

Khandkar, S. H. 2009. Open coding. *University of Calgary*, 23(2009): 2009.

Lalonde, D. 2021. Does cultural appropriation cause harm? *Politics, Groups, and Identities*, 9(2): 329–346.

Lee-Kan, S. M.; Bravo, N.; Hogan, C.; Filipowicz, A. L. S.; and Shamma, D. A. 2024. Assessing and Enhancing the Perceived Utility of Sustainable Driving Information.

Louie, R.; Coenen, A.; Huang, C. Z.; Terry, M.; and Cai, C. J. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020*

CHI conference on human factors in computing systems, 1–13.

Lyria Team; Caillon, A.; McWilliams, B.; Tarakajian, C.; Simon, I.; Manco, I.; Engel, J.; Constant, N.; Li, P.; Denk, T. I.; et al. 2025. Live Music Models. *arXiv preprint arXiv:2508.04651*.

Peng, Z.; Wang, Z.; and Deng, D. 2023. Near-duplicate sequence search at scale for large language model memorization evaluation. *Proceedings of the ACM on Management of Data*, 1(2): 1–18.

Riehmann, P.; Hanfler, M.; and Froehlich, B. 2005. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 233–240. IEEE.

Samuelson, P. 2023. Generative AI meets copyright. *Science*, 381(6654): 158–161.

Shan, S.; Cryan, J.; Wenger, E.; Zheng, H.; Hanocka, R.; and Zhao, B. Y. 2023. Glaze: Protecting artists from style mimicry by {Text-to-Image} models. In *32nd USENIX Security Symposium (USENIX Security 23)*, 2187–2204.

Somepalli, G.; Singla, V.; Goldblum, M.; Geiping, J.; and Goldstein, T. 2023. Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6048–6058.

Terry, G.; Hayfield, N.; Clarke, V.; Braun, V.; et al. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology*, 2(17-37): 25.

Thickstun, J.; Hall, D.; Donahue, C.; and Liang, P. 2024. Anticipatory music transformer. *Transactions on Machine Learning Research*.

UMG Recordings, Inc. v. Suno, Inc. 2024. (1:24-cv-11611). 2024; Main complaint filed 06/24/2024.

UMG Recordings, Inc. v. Uncharted Labs, Inc. 2024. (1:24-cv-04777) 2024; Main complaint filed 06/24/2024.

Vanka, S. S.; Steinmetz, C.; Rolland, J.-B.; Reiss, J.; and Fazekas, G. 2024. Diff-mst: Differentiable mixing style transfer. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.

Vollstedt, M.; and Rezat, S. 2019. An introduction to grounded theory with a special focus on axial coding and the coding paradigm. *Compendium for early career researchers in mathematics education*, 13(1): 81–100.

Wu, Y.; Chen, K.; Zhang, T.; Hui, Y.; Berg-Kirkpatrick, T.; and Dubnov, S. 2023. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Yang, X.; Guan, J.; Xiao, F.; Fan, C.; Liu, H.; Zhu, Q.; Xu, D.; and Lin, Y. 2025. DualMark: Identifying Model and Training Data Origins in Generated Audio. *arXiv preprint arXiv:2508.15521*.

Yin, Z.; Reuben, F.; Stepney, S.; and Collins, T. 2022. Measuring when a music generation algorithm copies too much: The originality report, cardinality score, and symbolic fingerprinting by geometric hashing. *SN Computer Science*, 3(5): 340.

Zhang, J. D.; and Schubert, E. 2019. A single item measure for identifying musician and nonmusician categories based on measures of musical sophistication. *Music Perception: An Interdisciplinary Journal*, 36(5): 457–467.

Zhang, Z.; Ning, Z.; Xu, C.; Tian, Y.; and Li, T. J.-J. 2023. PEANUT: A human-AI collaborative tool for annotating audio-visual data. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–18.

Zhao, Y.; Qiu, L.; Ai, W.; Shi, F.; and Zhu, S.-C. 2020. Vertical-Horizontal Structured Attention for Generating Music with Chords. *arXiv preprint arXiv:2011.09078*.

A Survey Questions

Below are the survey questions participants were asked about each of their generated songs. They were asked the same set of questions six times total, three times for the pre-tool phase of the interview, and three times for the post-tool phase of the interview. All multiple choice questions were on a 5-point scale.

- Please name the genre(s) (e.g., Indie folk, bossa nova, etc.) this excerpt seems most similar to. If the genre seems unrecognizable to you, put "I don't know."
- How confident are you in your answer to the previous question? ([Multiple Choice] Not confident at all - Extremely confident).
- Please name the artist(s) (e.g., Taylor Swift, Beethoven, etc.) whose music seems most similar to this excerpt. Feel free to include multiple artists if they come to mind. If none come to mind or you can't name them, put "I don't know."
- How confident are you in your answer to the previous question? ([Multiple Choice] Not confident at all - Extremely confident).
- How likely do you think someone that recognizes music by the artist you named would think this song is by that artist? ([Multiple Choice] Extremely unlikely - Extremely likely).
- How similar does this song seem to any existing songs or artists you are aware of? ([Multiple Choice] Not at all - Extremely Similar)
- Please describe how comfortable you would be doing the following activities with what you currently know about the song ([Multiple Choice] Not comfortable at all - Extremely comfortable):
  - Releasing this song on Spotify
  - Using this song in a vlog or YouTube video
  - Using this song in a podcast

B Post-Test Interview Questions

The following questions were asked during the post-test interview phase of the sessions. These averaged 5 minutes in length, and each participant was asked a subset of these questions relevant to their session.

General questions:

- Can you explain your thought process behind how your answers changed or did not change before you had access to the training data and afterwards?
- What went into your decisions about the final three questions on the survey—releasing on Spotify, using in a YouTube video, and using in a podcast—what affected your level of comfort with performing these activities with the generated song?
- How would you describe your confidence levels changing or not changing before and after having access to songs from the training data?
- Have you ever made music with generative music models before?
- If you were to use a generative music tool, would you prefer to have something like this where it exposed the training data? Or would you prefer to be more exploratory and not have access to the training data information?
- *(Follow up if they said they liked the tool)* What did you like about using the tool versus not having access to it?
- What made you more likely to click those genre links and explore the genres or the YouTube links—how did you find those useful or not useful, and what was your thought process behind those choices?

Musician specific questions:

- As a musician, would you see yourself using this in a work flow? Or is this outside of what you would normally use?
- When you're drawing inspiration from different artists and samples and sounds in general, how would you describe your process of knowing where songs come from versus your own novelty?
- What are your general opinions towards generative music as a musician? Do you feel as if it affects you, do you have any strong opinions towards it?

Media maker specific questions:

- What type of media do you prefer to make?
- What is your process for acquiring music or choosing music for [their type of media]? Or is that outside of your purview?
- Would you be open to using AI generated music in your work? Or do you have concerns about that?
- You were saying you would use a tool like this to [various uses in their workflows]. How would you use this tool to satisfy that need?

C Genre Identification

In the main text we analyzed changes in users' ability to identify artists before and after having access to *AI*-tribution. In this section, we perform the same analysis for genres. The full output of the categorized answers users gave can be seen in Figure 4. This flow figure displays the first answer users gave and what they changed (or did not change) their answer based on information provided via *AI*-tribution. A breakdown of this figure for the three user groups can be found in Appendix Figure 6.

For ***high similarity songs***, many participants first name a generic version of the genre on the tool (e.g., "jazz"), and then either maintain their answer, change it to a new answer the tool displayed, or become more descriptive in their answer. For those who had an answer outside of the top three songs prior to seeing outputs from *AI*-tribution or put "I don't know," they all changed their answer to something influenced by the interface (either something on the tool directly or discovered through exploring the linked genre pages).

For ***low similarity songs***, music consumers were consistent in not knowing the genre both before (80%) and after

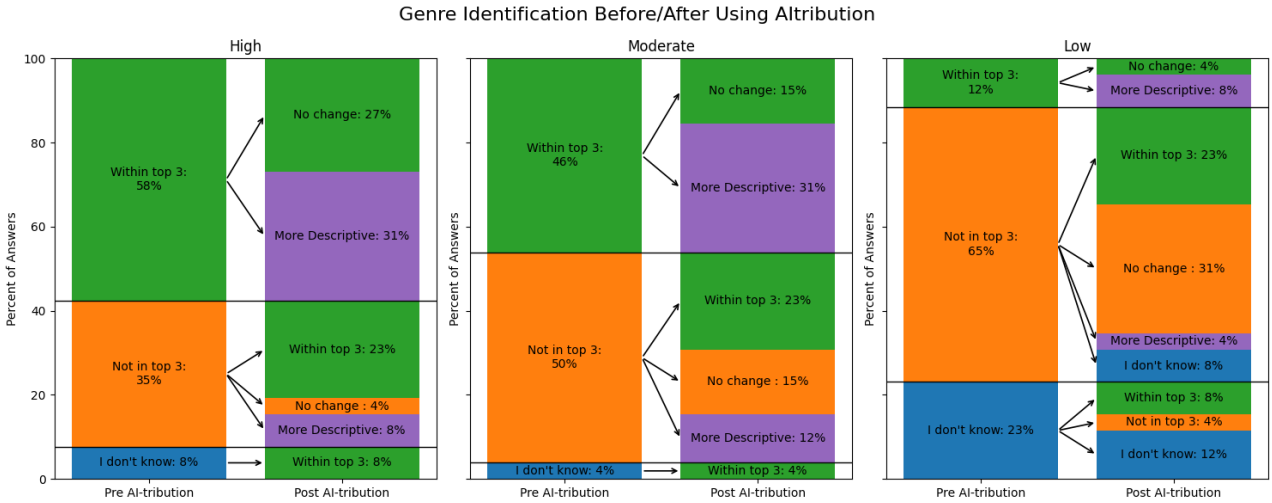

Figure 4: Variation of a Sankey diagram (Riehmann, Hanfler, and Froehlich 2005) displaying the distribution of answers given for genre identification and how they changed pre- (left bar) and post-access (right bar) to *AI*-tribution's top 3 songs in the training data. Split across high, moderate, and low behavior songs.

(70%) seeing the training data. Musicians were likely to either get more descriptive after seeing the suggestion, change their answer to something not displayed in the top 3 songs on the interface, or maintain that they do not know. One music consumer (P15) noted in these instances they "felt more certain about what it was not than what it was." Media makers on the other hand were likely to change their answer from something outside of the three top songs in the training data in the pre-phase (70%) or "I don't know" (20%) to something displayed in the top 3 songs (70%) in an act of blind faith. One media maker (P8) described this phenomenon: "I felt more confident that it probably knows what type of genre this is fitting in better than I do. Because these are not styles I'm familiar with." One music consumer (P11), shared this sentiment: "I never really understood how genre categorization works, so seeing [the outputs] made me believe it more, which genre it was supposed to be in. I would rather believe what the training data says it is." These participants would rather trust what the tool identified as the genre because they did not trust their own instinct.

Music consumers ascribe *AI*-tribution with providing them a new vocabulary to describe genres of songs. P7 remarks, "I didn't even have the right vocabulary to describe a lot of the music. I think my head goes mainly to 'pop,' 'jazz,' just very broad terms. So just to know more of the niche terms; I never would have come up with 'dutch pop' on my own if you left me to my own devices." One music consumer (P9) agreed with this idea: "I think the tool kind of gave me a vocabulary to think about things like genre...I remember seeing 'Russian post-punk' and that's a genre I didn't know existed, but that feels like a better fit than what I've been calling 'hard rock.' It was useful to have those nudges." Another music consumer (P24) even said that having access to better genre information fed into their ability to identify similar artists: "I am bad at putting exact words to a song genre. So I think it was helpful to give me words but also to connect artists that I had a hunch would be in this category."

Musicians tended to already have an extensive genre knowledge. But even they occasionally identified the tool as encouraging them to be more specific; P5 noted, "I think it gave me more of a repertoire; before I was just thinking of more generic names like 'Okay this is clearly jazz, this is more bluesy less bluesy, bouncy, and I would try to place them into these buckets. But when I had the tool I had a direction I could follow and when I opened [Every Noise at Once] I could see all those names and all those more specific genres. Then it was more easy to draw the line and it gave me more knowledge."

D Distribution of Music in Study

Distribution of Music Used in Our Study			
2***Similarity**	**Genre Subset**		
Behavior	**Total**	**Jazz/Blues**	**Pop/Orchestral**
Total	15	7	8
High	5	2	3
Moderate	5	2	3
Low	5	3	2

Table 3: Distribution of the songs used in our user study. We had 15 input songs total split across three categories of song behavior: high, moderate, and low similarity. We pulled input songs from two different subsets of VampNet's training data: a jazz/blues subset and a pop/orchestral subset.

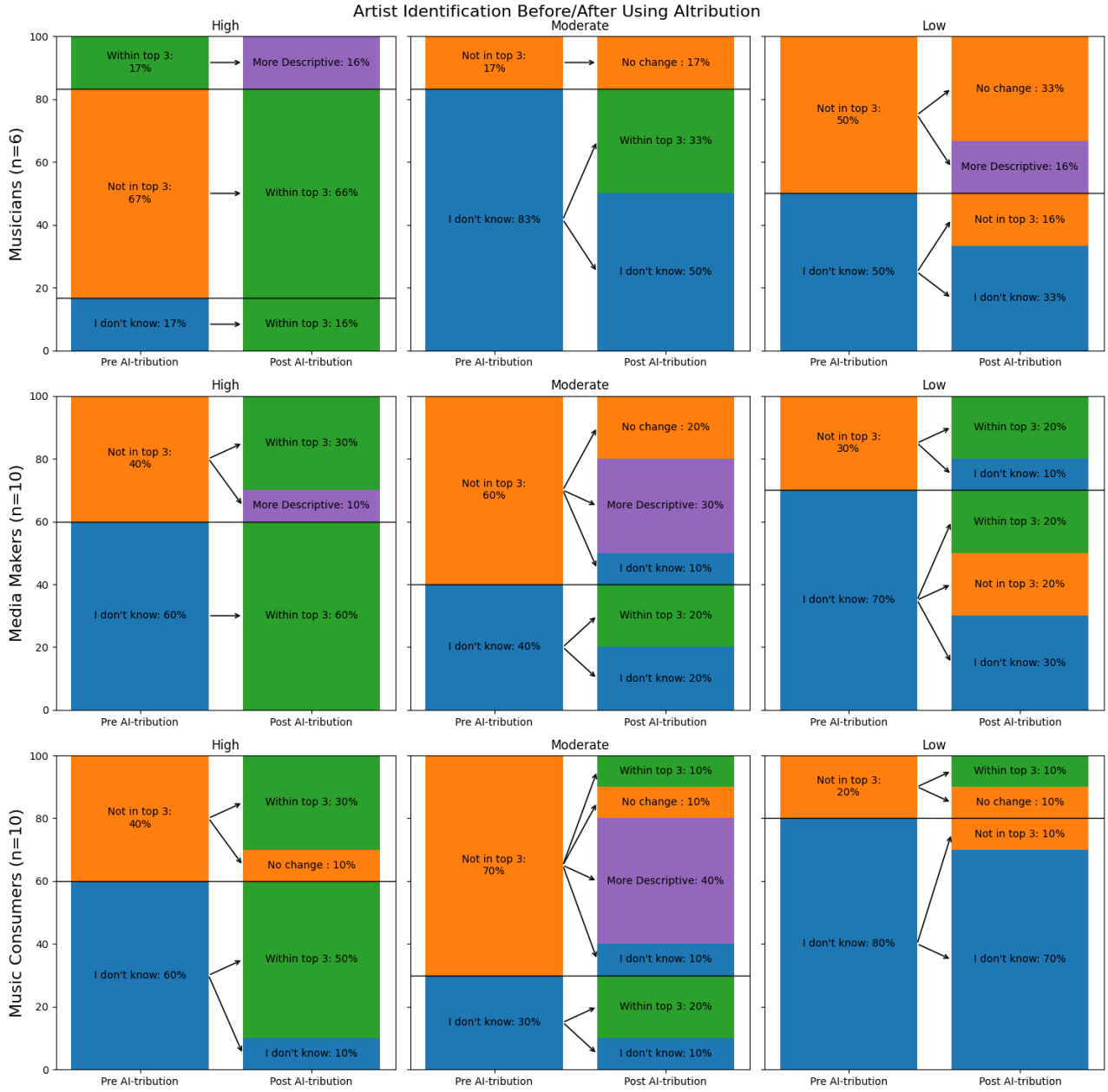

Figure 5: Plot displaying the distribution of answers given for genre identification and how they changed pre- (left bar) and post-access (right bar) to *AI*-tribution's top 3 songs in the training data. Split across low, moderate, and high behavior songs (in columns) and the three user groups (in rows).

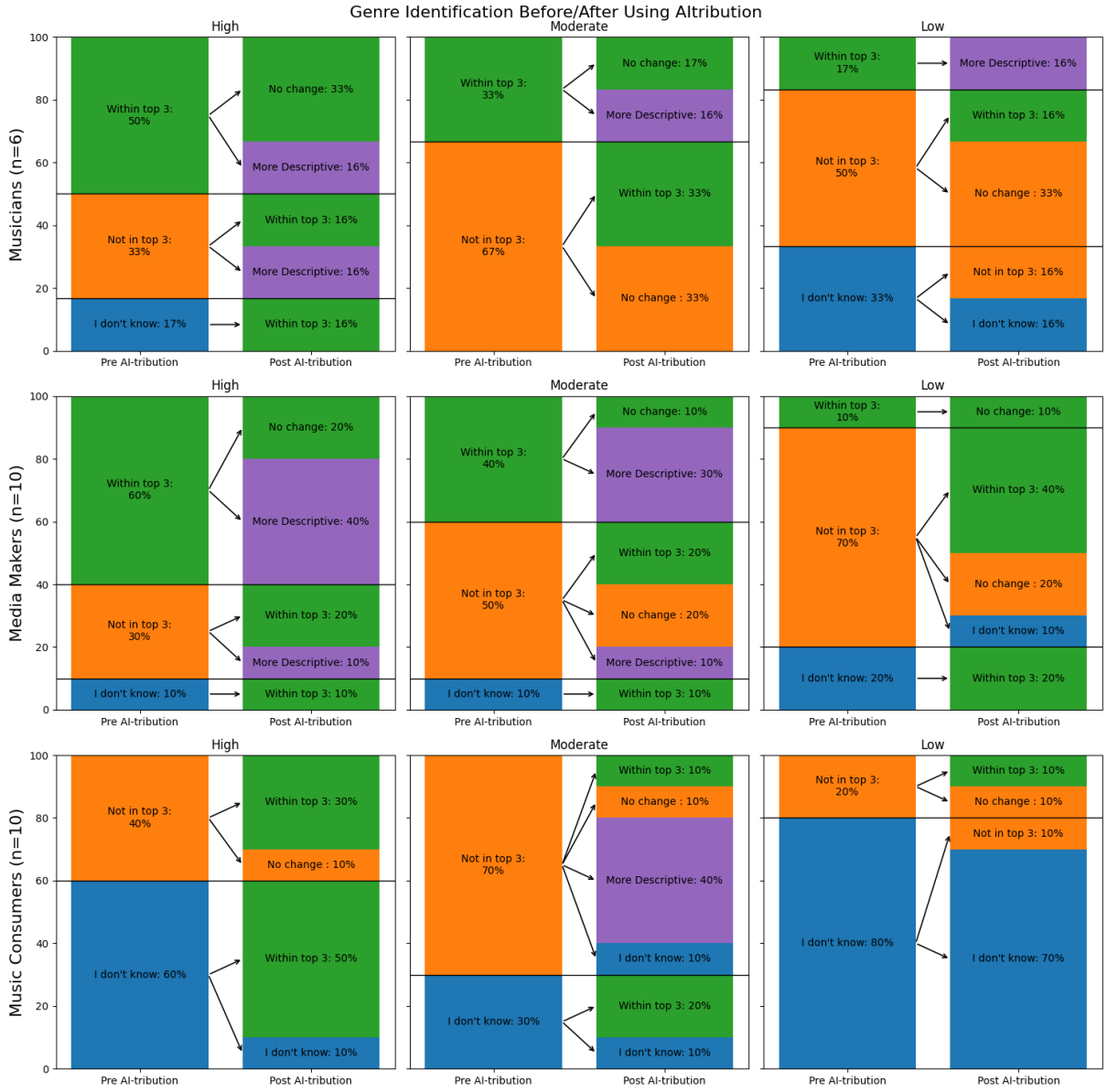

Figure 6: Plot displaying the distribution of answers given for genre identification and how they changed pre- (left bar) and post-access (right bar) to *AI*-tribution's top 3 songs in the training data. Split across high, moderate, and low behavior songs (in columns) and the three user groups (in rows).

E Change in Confidence, Identifiability, and Comfort Level by User Group

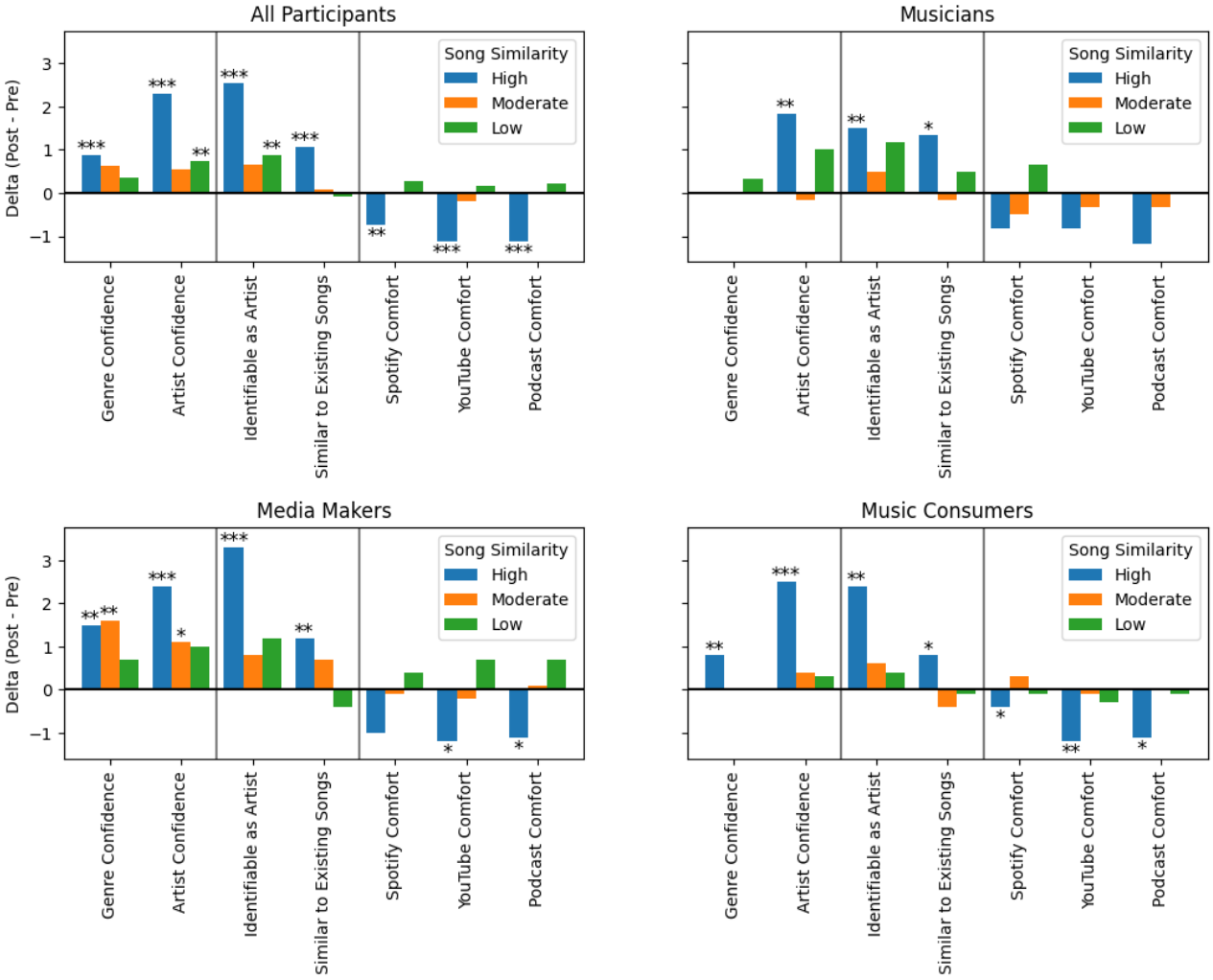

Figure 7: Bar plots displaying the changes between pre- and post-access (calculated as post value - pre value) to *AI*-tribution information in confidence scores, degree of similarity evaluations, and comfort levels for using the generated songs in various use-cases. Split by subplot for all participants, musicians, media makers, and music consumers. Bar colors from left to right indicate the change for high-, moderate-, and low-similarity songs. Positive bars represent increases in e.g., confidence or comfort, and negative bars represent decreases in e.g., confidence or comfort. Asterisks above/below the bars indicate significance scores for paired sample t-tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

F Media Makers' Attitudes Towards Music

The quantitative analysis revealed that media makers were the most likely of the three user groups to have blind faith in the top three similar songs identified by *AI*-tribution and not actually use the resources at their disposal unless guided; the qualitative analysis reveals why. The first reason is that media makers tended to have self professed low music literacy. One (P7) said so in those exact terms: "I think I have very low music literacy and I wish that wasn't the case but it is." Another (P14) described an effect of that while detailing what would drive their comfort in using generated music: "I think mostly if the music sounds pretty generic to me, that would make me more comfortable using it for my content."

The second, and seemingly more resonant, reason is there was a shared sentiment by many media makers that music came as an afterthought to whatever their primary medium of content was. P10 described their workflow for producing podcasts: "Music was always the add-on extra task, which I had no training for, and it'd really lift an episode. But if you don't know anything about music, then it was the last thing you did." P7 described a similar mentality in their own workflow:"...when I think about creating videos or something I think of music almost as an afterthought in my media."

Media makers consistently discussed using the tool to prevent their own content to get flagged for copyright violations. P2 remarked, "This is a must have for if I'm trying to use AI generated music as a source in my video or something. Just because...I don't want my comment section to be filled with 'You're copying music from others!' so I want to be prepared." This was a sentiment shared by others. Media makers' complex attitude towards music as a peripheral activity to their main medium drives them to have a greater dependence on attribution tools than either musicians or music consumers. They saw this information as almost a prerequisite to the use of AI tools, and expressed their desire for this information as more of a need than a preference.