COMPUTATIONAL COPYRIGHT: TOWARDS A ROYALTY MODEL FOR AI MUSIC GENERATION PLATFORMS

Junwei Deng, Jiaqi Ma
School of Information Sciences
University of Illinois Urbana-Champaign
{junweid2,jiaqima}@illinois.edu

ABSTRACT

The advancement of generative AI has given rise to pressing copyright challenges, particularly in music industry. This paper focuses on the economic aspects of these challenges, emphasizing that the economic impact constitutes a central issue in the copyright arena. The complexity of the black-box generative AI technologies not only suggests but necessitates algorithmic solutions. However, such solutions have been largely missing, leading to regulatory challenges in this landscape. Focusing on the music domain, we aim to bridge the gap in current approaches by proposing potential royalty models for revenue sharing on AI music generation platforms. Our methodology involves a detailed analysis of existing royalty models in platforms like Spotify and YouTube, and adapting these to the unique context of AI-generated music. A significant challenge we address is the attribution of AI-generated music to influential copyrighted content in the training data. To this end, we present algorithmic solutions employing data attribution techniques. Our experimental results verify the effectiveness of these solutions. This research represents a pioneering effort in integrating technical advancements with economic and legal considerations in the field of generative AI, offering a computational copyright solution for the challenges posed by the opaque nature of AI technologies.

1 INTRODUCTION

Recent advancements in generative AI have significantly impacted creative industries, leading to a surge in AI-generated content across art, music, literature, and software. This rapid evolution has raised complex legal challenges, especially concerning copyright issues (Henderson et al., 2023; Samuelson, 2023; Sag, 2023; Franceschelli & Musolesi, 2022). A notable instance of these challenges is the recent lawsuit filed by New York Times against Microsoft and OpenAI (NYT, 2023). Copyright laws cover a range of rights, including protection of original works, controlling their reproduction, and managing the distribution of profits from these works. The emergence of generative AI poses multifaceted challenges in this regard, as it blurs the lines of authorship and originality.

Arguably, central to these challenges is the economic impact. Taking the music industry as an example, a vast collection of music has been publicly available on platforms like Spotify and YouTube, where copyright owners are compensated through royalties. This practice not only suggests that economic incentives are a primary reason for making music publicly accessible, but also highlights the centrality of economic rights in copyright protections. This trend is reflective of a broader truth: economic considerations are at the heart of the U.S. copyright law, where a primary goal is to stimulate creativity by ensuring that creators are adequately compensated. There has also been ongoing debate about whether training generative AI with copyrighted content aligns with the *fair use* doctrine[1]. However, it is increasingly argued that fair use may not apply if the AI generated content competes with the original market for the data (Henderson et al., 2023). These issues underscore the economic impact as a crucial aspect of copyright challenges in generative AI.

[1]See Section 107 of the Copyright Act: https://www.copyright.gov/title17/92chap1.html#107.

However, effective technical solutions addressing the aforementioned challenge have been limited or nonexistent. Existing efforts have focused on preventing generative AI from generating content similar to its training data (Vyas et al., 2023; Chu et al., 2023; Li et al., 2023). This approach, while helpful, may not fully address the broader economic implications of AI-generated content. Addressing the economic aspect of the copyright challenges is challenging as it requires a solution that integrates technical advancement into business agreements.

The challenge is also pressing. Without effective technical solutions for proper royalty distribution, regulatory bodies are faced with a dilemma between stifling innovation and compromising the interests of copyright owners. As it stands, numerous AI music generation platforms are navigating these uncharted waters, operating in legal gray areas and leaving the rights of copyright owners inadequately protected (Drott, 2021; Clancy, 2021).

This paper aims to bridge this crucial gap by proposing potential royalty models for revenue sharing from AI music generation platforms. Specifically, we design the royalty model by addressing the following key questions: 1) Who are the stakeholders? 2) What are the sources of revenue? 3) How to determine the royalty distribution for revenue sharing? To answer these questions, we start with case studies of Spotify and YouTube, which are the leading platforms in music streaming and video sharing respectively. We investigate their royalty models and examine feasibility of adapting these models to AI music generation platforms. A critical technical challenge for such adaptation we identify is the difficulty in attributing the AI generated music to the influential copyrighted content used in the model training data. In response, we develop algorithmic solutions using data attribution techniques to mitigate these challenges. Our experimental results demonstrate that the proposed solutions are reasonably effective.

The proposed approach represents an early effort to navigate the complex intersection of technological innovation and economic considerations in copyright law for generative AI. The complexity of the black-box generative AI technologies necessitates a computational copyright solution. This paper showcases a promising prototype towards this goal.

2 Exploring Digital Music Royalty Models Through Case Studies

In this section, we examine the royalty models in the digital music industry through a couple of case studies. Please refer to Appendix A for fundamental concepts and a few major types of music royalties that are prevalent in the industry. In order to understand the intricacies of the implementation of royalty models and their applicability to AI music generation platforms, we delve into case studies of two major platforms: Spotify and YouTube. Spotify is the largest music streaming platform in the world while YouTube is the largest video sharing platform. Both platforms have a significant amount of music content and generate revenue through multiple sources. Furthermore, despite various existing criticisms on these royalty models (Marshall, 2015; Trendacosta, 2020), they represent the status quo of how the current digital music market works. Therefore, designing royalty models for AI music generation platforms mimicking the ones for Spotify and YouTube would be a reasonable initial step in this area. In the following sections, we will examine the royalty models of these two platforms in detail.

2.1 Case Study: Spotify

Spotify employs a centralized method for sharing its revenue with copyright owners, primarily via *streaming royalties*. The process involves determining Spotify's total revenue from various sources and subsequently calculating the royalty distribution for copyright owners.

Stakeholders. Spotify's royalty model involves several key groups of stakeholders, in addition to the streaming platform itself. These groups[2] are (1) *artists and creators*, (2) *record labels and music publishers*, (3) *music rights societies and collecting agencies*, (4) *listeners and subscribers*, and (5) *advertisers*. Stakeholders in groups 1, 2, and 3 receive revenue shares from Spotify, while groups 4 and 5 contribute to the generation of Spotify's revenue. Typically, Spotify directly interacts with stakeholders in groups 2 and 3. Individual artists and creators often have contracts with these labels,

[2] Please refer to Appendix B for detailed description of these groups of stakeholders.

publishers, or music rights agencies, and do not directly engage with Spotify in the financial aspect of their music streaming.

Revenue Sources. The major revenue sources of Spotify can be divided into two categories: subscription and advertisement. In 2021, premium subscriptions accounted for 88% of Spotify's revenue while advertisements accounted for the remaining 12% (Johnston, 2023). The two revenue sources lead to the formation of separate revenue pools, which are also calculated separately for different countries or regions.

Royalty Distribution. Spotify employs a straightforward *pro rata* model to calculate the royalty distribution for each revenue pool. The royalty for each artist or track is calculated by applying their stream share to each revenue pool. This method ensures that royalty distribution is directly proportional to the popularity and streaming frequency of each artist's or track's work on the platform.

2.2 Case Study: YouTube

YouTube's model for compensating music copyright owners is multifaceted, offering various methods for monetizing the content:

1. **YouTube Partner Program:** Music copyright owners can join the YouTube Partner Program, uploading music (videos) to their official channels. Revenue is shared based on user views of their content.
2. **Content ID:** Owners can earn from videos using their music through the *Content ID* system. This system uses fingerprinting and machine learning to identify copyrighted content in uploaded videos and allocates revenue from these videos to the copyright owners.
3. **Licensing:** Owners can also license their music directly to a YouTube video for a one-time payment.

The first method resembles Spotify's royalty model. The second and the third methods are different as they involve a third-party video creator.

Stakeholders. The stakeholders involved in the first method above are similar to those in Spotify's royalty model. However, the second and third methods introduce additional parties: video creators and third-party licensing platforms. Video creators are the ones who upload videos (incorporating copyrighted music) to YouTube. Third-party licensing platforms are companies that help video creators obtain licenses for music used in their videos. These companies often have direct licensing agreements with YouTube and music rights owners, offering a streamlined process for video creators to legally use music in their videos.

Revenue Sources. For the first two methods, royalties come from YouTube's revenue streams. YouTube generates revenue primarily through advertisements and, to a lesser extent, through premium subscriptions. The advertisement model is diverse, including in-video ads, banner ads, and sponsored content. Premium subscriptions, offering ad-free viewing and other benefits, also contribute to YouTube's revenue.

Royalty Distribution. In the first two methods, YouTube first calculates the revenue share for each video in a manner similar to Spotify. However the calculation is more complex due to the diverse range of advertisement models and varying user interactions with the videos. For the second method, the revenue share for each video is further distributed to both video creators and music copyright owners, according to the Content ID system. This distribution can be automated, using an algorithm that assesses the actual use of music in the video, or it can be managed through manually negotiated agreements. In contrast, the third method employs a more straightforward royalty calculation, based directly on the licensing agreement.

A crucial aspect of YouTube's royalty model is the challenge of attributing copyright ownership in the videos. Unlike Spotify, where content attribution is straightforward through stream counts, the incorporation of copyrighted music in the user-generated videos makes this task technically demanding. The Content ID system serves as the technical foundation that enables the second and third methods of YouTube's revenue sharing. In the second method, it identifies music in videos and allocates revenue to the copyright owners. In the third method, while synchronization licenses might be obtained through third-party licensing platforms outside of YouTube, the presence of the Content ID system encourages them to secure these licenses. Although the system has its share of flaws and has faced criticism (Van der Sar, 2021; McKay, 2011; Trendacosta, 2020; Saadatpanah et al., 2020), including issues with false positives and negatives, it is still broadly embraced by the music industry.

2.3 SUMMARY

A notable pattern of the royalty models on digital music/video platforms is that the payment is not directly made to each individual piece of music work[3]. Rather, the platforms accumulate overall revenue from their services involving a vast collection of music, and then distribute a portion of the revenue among the copyright owners.

The typical methods for determining how revenue is distributed can be summarized in two main steps:

1. **Formation of Revenue Pools:** Determine the total amount of revenue available for distribution in each revenue pool. These pools are defined by various criteria, including the sources of revenue and geographic regions.
2. **Distribution Based on Access Frequency:** Within each revenue pool, distribute the revenue proportional to the frequency that each piece of music is played or accessed.

For example, on Spotify, revenue pools are based on subscription and advertisement; the revenue distribution is determined by the stream counts. On YouTube, the step 2 needs to be done in two sub-steps, where one first determines the frequency of a video is played, then attributes the copyright ownership through the Content ID system.

3 POTENTIAL ROYALTY MODELS FOR AI MUSIC GENERATION PLATFORMS

This section explores potential royalty models for AI music generation platforms. We start by understanding the business models of these platforms, which involves summarizing their services, identifying key stakeholders, and highlighting various revenue sources integral to their operations. This foundation sets the stage for discussing the design of proper royalty distribution mechanisms.

3.1 THE BUSINESS MODELS

While the landscape of AI music generation is still rapidly evolving, there have been a few common business models emerging (McFarland, 2023). We summarize these business models in terms of services, stakeholders, and revenue sources[4].

The backbone of AI music generation platforms is generative AI trained on a large corpus of existing music, which often includes copyrighted music. With the generative AI, the platforms offer a variety of services to meet different needs of end users. The potential stakeholders involved in AI music generation platforms have significant overlaps with those on traditional music platforms, as summarized in the five groups in Section 2.1. The platforms have several different ways for generating revenues, such as subscription fees, licensing, advertisements and costom composition fees.

3.2 POTENTIAL ROYALTY MODEL DESIGNS

Given the similarity of the stakeholders and revenue sources between the AI music generation platforms and traditional music platforms, it is logical to consider adopting and adapting existing royalty models from platforms like Spotify and YouTube.

Particularly, the business models for AI music generation platforms align with the pattern identified in Section 2.3, where the revenue is channeled through the platform, rather than directly compensating for each individual piece of music.

1. **Formation of Revenue Pools:** Similar to Spotify and YouTube, these platforms would first accumulate revenue, forming distinct pools based on different criteria such as revenue sources (subscriptions, advertisements, licensing fees).
2. **Distribution Based on Access Frequency:** The revenue from each pool would then be distributed based on the frequency at which each copyrighted work included in the training corpus is accessed during the service.

[3] The licensing model on YouTube, where video creators directly buy synchronization licenses for individual music, is an exception. However, note that this model is also enabled by the existence of the Content ID system.

[4] Please refer to Appendix C for more detailed AI music generation platforms business models.

The key question is how the copyrighted training content is "accessed" in the services provided by the platforms. Here, the music generated by a generative AI is influenced by the copyrighted works included in its training corpus. This scenario is analogous to YouTube, where copyrighted music is used as ingredients for new creations like videos or remixes. In the generative AI scenario, end users can be viewed to access the copyrighted training content indirectly through the generated music.

Recalling YouTube's model, the first step involves calculating the frequency of video views. Subsequently, these views are attributed to the copyrighted music used in the videos. For AI music generation platforms, a similar method could be employed: first determining the usage frequency of the generated music and then attributing this usage back to the original copyrighted works that influenced the creation of this music.

Enforcing such a royalty model presents an open technical challenge: accurately attributing the influence of original copyrighted works on the generated music. However, if this attribution challenge can be effectively addressed, the remaining elements of the royalty model can closely mirror those of YouTube's Content ID system.

In the following section, we propose an algorithmic solution to mitigate this challenge of attribution, aiming to create an effective royalty model for AI music generation platforms.

4 Attributing AI-Generated Music to Copyrighted Content

4.1 Formulation of the Attribution Problem

Attributing the influence of copyrighted training content on generated music essentially asks the question: "To what extent does each piece of music in the training corpus influence a specific piece of AI-generated music?" The quantifiable definition of "influence" can potentially be subjective To this end, we suggest two perspectives to define the "influence", one inspired by the machine learning literature and the other comes from the domain knowledge of music.

Perspective 1: Data Attribution. The *data attribution* problem in machine learning refers to the task of identifying and quantifying the contribution of each data point in the training dataset to a given output of a machine learning model. Formally, this problem is often framed as follows: How does the removal of a particular data point from the training dataset and subsequent retraining of the model affect its output? This change in output serves as a measure of the removed data point's influence on that specific model output (Koh & Liang, 2017). In the context of AI music generation, we can define the influence of a piece of training music on a piece of generated music in terms of the change in the likelihood of the model producing that generated music, assuming the model is retrained after removing the training music piece.

Perspective 2: Musical Influence. The second perspective considers the influence from a musical standpoint, focusing on how one musician's work might affect another's. Such influence spans multiple aspects, including musical styles (such as genres, rhythms, melodies, or harmonies), technical and instrumental methods (how a musician plays an instrument or sings), or thematic elements (such as themes, messages, or lyrical content).

In Section 4.2, we introduce an algorithm designed to estimate influence from the data attribution perspective. Then, in Section 4.3, we evaluate the proposed method using metrics from both perspectives, highlighting a potential synergy between these two viewpoints.

4.2 Data Attribution for Music Generation

In this section, we first introduce formal definitions of AI music generation and the data attribution problem for music generation. Subsequently, we propose an algorithmic solution designed to quantitatively estimate the influence of each training piece on a particular piece of generated music.

4.2.1 AI Music Generation

In the field of AI music generation, there are two major paradigms: *waveform music generation* and *symbolic music generation* (Manzelli et al., 2018). Waveform music generation involves the direct synthesis of a music's waveform, with examples including WaveNet (Oord et al., 2016). On the other hand, symbolic music generation involves creating music in a symbolic format, such as the Musical Instrument Digital Interface (MIDI) format. This paper focuses on symbolic music generation.

Symbolic Music Generation We start by introducing the notations and the formal definitions of symbolic music generation. Symbolic music is characterized as a series of discrete events that collectively represent a segment of music. A music segment m with k events can be expressed as $(e_1, e_2, \ldots, e_k)$. Here, each $e_k \in \mathcal{V}$ represents a specific event from a vocabulary set $\mathcal{S}$ that defines all the valid symbolic events.

A symbolic music generation model, denoted as h, takes a prompt music segment and calculates the subsequent event's conditional probability distribution as follows:

$$P(e_k \mid e_{k-p}, e_{k-p+1}, \ldots, e_{k-1}; h) = h(e_{k-p}, e_{k-p+1}, \ldots, e_{k-1}), \tag{1}$$

where $P(e_k \mid e_{k-p}, e_{k-p+1}, \ldots, e_{k-1}; h)$ is the conditional probability distribution of an event e_k, and p indicates the length of prompt. This type of model formulation is known as an autoregressive model, which is one of the most popular model family for symbolic music. This approach sequentially predicts the probability of new events based on the preceding events, and then selects an event from these probabilities.

In this paper, we ground our study with Music Transformer (Huang et al., 2018), one of the most popular autoregressive symbolic music generation models that employs the Transformer architecture (Vaswani et al., 2017) for the generation model h. Suppose the size of the vocabulary set is $|\mathcal{V}| = V$. The model h can be represented as a neural network with V output units, each of which corresponds to a specific event in the vocabulary set, and the V output units sum up to one. The output of the model corresponds to a probability distribution over the vocabulary set, and the model h can be viewed as a classifier with V classes. The model is trained by maximizing the log-likelihood over a training corpus of music.

4.2.2 Data Attribution for Symbolic Music Generation

Attribution Scores. Now we formalize the data attribution problem for symbolic music generation. Suppose we have a training dataset $S = \{m_1, m_2, \ldots, m_n\}$ with n segments of music and a generation model h_S trained on this dataset. For any piece of music segment m and any model h, we define a utility function $f(m, h)$ that maps the music segment m and the model h to a real value. The influence of a training piece $m_i \in S$ ($i = 1, \ldots, n$) on a new piece of AI-generated music $\hat{m}$, which is also called an *attribution score*, can then be defined as

$$I(m_i, \hat{m}) = f(\hat{m}, h_{S \setminus \{m_i\}}) - f(\hat{m}, h_S), \tag{2}$$

where $h_{S \setminus \{m_i\}}$ is the model retrained on the dataset S with m_i removed. In practice, the utility function $f(m, h)$ can be defined as the (log-)likelihood of music segment m being generated by model h. In this case, $I(m_i, \hat{m})$ measures the change of likelihood for $\hat{m}$ being generated when m_i is removed from the training corpus.

Two Levels of Attribution. We can define two instances of the data attribution problem, respectively *event-level attribution* and *segment-level attribution*. The event-level attribution corresponds to a special case where $\hat{m}$ has a single event, i.e., $|\hat{m}| = 1$. The segment-level attribution corresponds to the general case where $\hat{m}$ has multiple events. The two instances provide different granularity of attribution scores. In an autoregressive symbolic music generation model, the music is generated event by event. Therefore, the training data points could have different influences when generating different events in a segment. The event-level attribution provides a way to capture this nuance. On the other hand, the segment-level attribution looks at the influence of training data on a larger scale, focusing on the overall structure and composition of a generated music segment.

Estimating the Attribution Scores. Directly calculating $I(m_i, \hat{m})$ requires retraining a model for for each training data point m_i, which is computationally prohibitive. Fortunately, there has been a rich literature on efficient data attribution methods (Hammoudeh & Lowd, 2022), primarily designed for classification models. Furthermore, we shall see that these methods can be easily adapted to the autoregressive generative models like Music Transformer.

In particular, since generating one event can be viewed as a classification problem, we can directly apply existing data attribution methods for classification models to event-level attribution. For the segment-level attribution, when the utility function is defined as the log-likelihood, i.e.,

$$f(\hat{m}, h) = \log P(\hat{m}; h)$$

we observe that, assuming $\hat{m} = (e_k, \ldots, e_l)$ for some $k < l$, then by Bayes' rule,

$$f(\hat{m}, h) = \log P(\hat{m}; h) = \sum_{i=k}^{l} \log P(e_i \mid e_{i-p}, \ldots, e_{i-1}; h) = \sum_{i=k}^{l} f(\{e_i\}, h).$$

This implies that the attribution score $I(m_i, \hat{m})$ can be decomposed as

$$I(m_i, \hat{m}) = f(\hat{m}, h_{S\setminus\{m_i\}}) - f(\hat{m}, h_S) = \sum_{i=k}^{l} \left(f(\{e_i\}, h_{S\setminus\{m_i\}}) - f(\{e_i\}, h_S) \right) = \sum_{i=k}^{l} I(m_i, \{e_i\}),$$

which is exactly the sum of all the event-level attribution scores. Therefore, we can apply any data attribution method that can attribute the log-likelihood of a classification model in the segment-level attribution when using log-likelihood as the utility function.

There are several off-the-shelf data attribution methods that can be applied to estimate the attribution scores for the autoregressive symbolic music generation models, such as Influence function (Koh & Liang, 2017), TraceIn (Pruthi et al., 2020), and TRAK (Park et al., 2023). In our study, we choose to use TRAK for its efficiency and state-of-the-art performance. We denote the estimated attribution score for m_i on $\hat{m}$ as $\hat{I}(m_i, \hat{m})$.

4.3 Experimental Evaluation

We experiment on Music Transformer (Huang et al., 2018) model and MAESTRO dataset (Hawthorne et al., 2019). For the implementation of attribution score calculation, we utilize the TRAK Park et al. (2023) toolkit as well as a random score algorithm as reference for both level attribution. Please refer to Appendix D for detailed experimental setup.

4.3.1 Evaluation Metrics

We introduce evaluation metrics formalizing the two perspectives of "influence" in Section 4.1.

Retraining Rank Correlation. To evaluate the estimated attribution scores from the data attribution perspective, we can compare them with the ground truth attribution scores defined in Eq. (2). In the data attribution literature (Koh & Liang, 2017; Ilyas et al., 2022), the comparison is typically measured by Spearman's rank correlation. Formally, for a training dataset with n data points, one will calculate the rank correlation between $\{\hat{I}(m_i, \hat{m})\}_{i=1}^n$ and $\{I(m_i, \hat{m})\}_{i=1}^n$.

However, calculating $\{I(m_i, \hat{m})\}_{i=1}^n$ involves retraining a model for removing each data point, which becomes computationally impractical on large datasets. Following Ilyas et al. (2022), we adopt an approximated version of this rank correlation metric. Instead of retraining for removing each data point, we randomly select a set of subsets of the training dataset, $U \subseteq 2^S$, and retrain a model on $S \setminus S'$ for each $S' \in U$. Slightly overloading the notation, we define a subset attribution score as $I(S', \hat{m}) := f(\hat{m}, h_{S\setminus S'}) - f(\hat{m}, h_S)$. Correspondingly, we use the summation of the estimated attribution scores on each subset as the estimated attribution score for that whole subset, i.e., $\hat{I}(S', \hat{m}) := \sum_{m\in S'} \hat{I}(m, \hat{m})$. Then we can calculate a rank correlation between $\{\hat{I}(S', \hat{m})\}_{S'\in U}$ and $\{I(S', \hat{m})\}_{S'\in U}$.

Musical Similarity. For the musical influence perspective, there are multiple aspects mentioned in Section 4.1. In our study, we focus on the similarity of musical styles. A common approach to quantitatively evaluate musical style similarity is by extracting features from the music (Slaney et al., 2008). In this study, we identify three features used in Spotify APIs [5] to characterize a piece of music. *Loudness* measures the overall velocity of a music segment. We define it as the average velocity of events within the segment. *Key* measures the average pitch height of all events in the music segment. *Duration* measures the total length of the music segment in time, calculated as the sum of the time deltas of all events.

We extract these features from both the generated music and the training samples. Then we can evaluate the attribution methods by investigating if the most influential training music pieces are more similar to the generated music in terms of musical styles. Formally, for each musical style feature, we calculate the Pearson correlation over pairs of generated music and training music pieces.

[5] `https://developer.spotify.com/documentation/web-api/`

4.3.2 Experimental Results

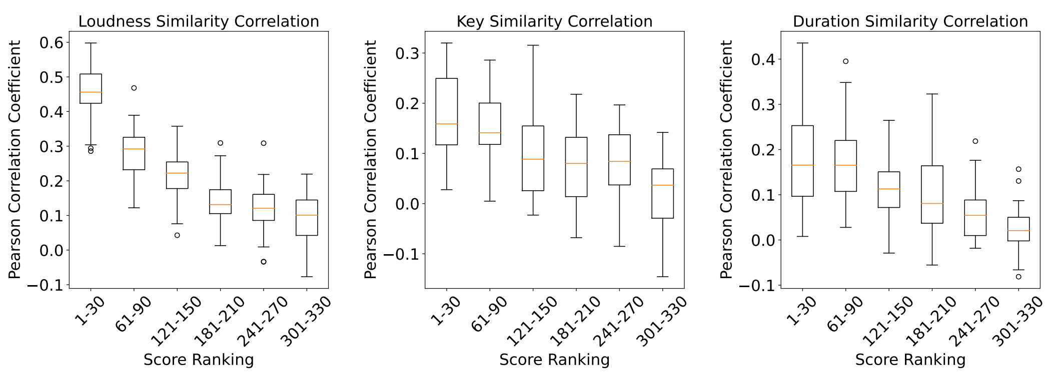
Retraining Rank Correlation. We form the set U with 100 random subsets, each contains 50% of the training samples. We calculate the rank correlations on 178 generated music and report the average rank correlations for different data attribution methods in Table 1. "Random score" refers to assigning a random score for each training sample. "TRAK (segment-level)" and "TRAK (event-level)" refers to different levels of attribution. In comprison to the random score baseline, both levels of attribution methods have achieved significantly positive correlations with the ground-truth scores. This indicates that there exist computationally feasible solutions that can reasonably attribute the generated music to the copyrighted training content, thus solving the key technical bottleneck for establishing a royalty model. In addition, we observe that event-level attribution seems to be easier than segment-level attribution. This leads to an interesting question about the proper granularity of attributing generated music, which we leave for future exploration.

Table 1: The average Spearman's rank correlation among 178 generated music for different data attribution methods.

Data attribution method	Random score	TRAK (segment-level)	TRAK (event-level)
Spearman's rank correlation	0.0091	0.301	0.359

Musical Similarity. For each generated music, we order the training music pieces by their attribution scores. Figure 1 shows the results of musical similarity between the generated music and training music pieces being decreasingly influential on the generated music. We observe a clearly decreasing trend of similarity in these plots, which suggests that the data attribution methods also capture some influence in terms of musical styles.

Figure 1: We calculated 178 pairs of generated music and the training sample with different ranks of the attribution score's loudness, key, and duration. The boxplots represent the statistics of the Pearson correlation coefficient for different rank ranges and are plotted for each feature.



5 Related Work

In this section, we discuss how the computational copyright solution and AI music royalty model in this paper are connected to prior works. These works can fall into three categories: (1) *law*, (2) *economy*, and (3) *technology*.

Law. Legal issues related to generative AI copyright have attracted considerable interest. One of the primary discussions is the applicability of the fair use doctrine. Though the fair use doctrine was shown to be applicable in data mining through previous studies (Sag, 2018) and lawsuits[6], recent studies continue to raise questions about its suitability for AI-generated content (Henderson et al., 2023; Samuelson, 2023; Sag, 2023; Franceschelli & Musolesi, 2022; Peukert & Windisch, 2023). One of the main reasons for this transformation is that AI-generated content could be substantially similar to original works and impact the market (Henderson et al., 2023; Peukert & Windisch, 2023). For additional discussions related to fair use, please see Appendix E.

[6]See `https://casetext.com/case/guild-v-google-inc-1`

Another challenge that is widely discussed is the test (Lee et al., 2023) to assert copyright infringement. The derivative works, like AI generation, may infringe on adaptation rights[7], as long as proper tests are satisfied. It is suggested that the tests for copyright infringement can be divided into two main categories: *access* and *similarity* (Vyas et al., 2023; Lee et al., 2023). Access considers whether any part of an AI generation was directly influenced by a specific training example, and similarity considers if the generation is substantially similar to a training example. Correspondingly, we provide two perspectives, i.e., training data attribution and similarity, to define "influence" in Section 4.1 and show the potential synergy of these two perspectives. The legal community has recognized a technical gap in the process of determining which training samples influence the generation (Lim, 2023; Lee et al., 2023; Gans, 2024). This paper represents a pioneering effort aimed at bridging this gap.

Economy. The effect of technical advancements, such as the digitalization of music, on the economic aspect of copyright has been discussed in numerous studies (Peukert & Windisch, 2023; Valdovinos Kaye et al., 2021; Gans, 2024). These discussions mainly focus on the social welfare of different copyright regimes. One of the most important regimes is *Notice and Takedown* (N&T)[8]. Systems following this regime are often employed by large platforms like YouTube (Van der Sar, 2021) and TikTok (Valdovinos Kaye et al., 2021). However, studies argue that these systems do not come without problems (Peukert & Windisch, 2023; Gans, 2015). Specifically, copyright owners, especially in the music industry (Gans, 2015), could benefit from derivative works like user-generated videos, but the forced taking down of derivative works undermines the welfare and promotion of copyright owners (Peukert & Windisch, 2023).

There is a trend that copyright owners are transitioning from using N&T systems to *algorithmic licensing* (Peukert & Windisch, 2023). Algorithmic licensing provides opportunities for creating a market where copyright owners could benefit from derivative works. For example, Content ID on YouTube provides monetizing option other than taking down for copyright owners. Empirical evidence indicates the monetizing option is widely chosen with a low dispute rate (cop, 2021). Remarkably, the royalty model and prototype proposed in Section 3 and Section 4 can be seen as an algorithmic licensing model for AI-generated music.

Technology. There have been several studies of technical solutions in the copyright arena, while most of the existing literature focuses on preventing generative AI from generating infringing works (Vyas et al., 2023; Chu et al., 2023; Li et al., 2023; Somepalli et al., 2023b;a). Li et al. (2023); Somepalli et al. (2023b;a) investigate the memorization effect, which refers to the generation of something directly copied from training data, and the mitigations like deduplication in the diffusion model. Vyas et al. (2023); Chu et al. (2023) propose protection methods based on algorithm stability to prevent models from generating very similar works to copyrighted input. Though there have been promising results in this area, studies (Elkin-Koren et al., 2023) show that these methods could have some downsides that deviate from the purpose of copyright law.

Another closely relevant direction is *data valuation*, which usually assigns test sample agnostic value scores to training samples (Ghorbani & Zou, 2019). However, the model for music royalty necessitates a more fine-grained approach to attribution, extending to each test sample or, in more detailed instances, to each individual event.

6 Conclusion

In conclusion, this paper has explored the intricate landscape of copyright challenges posed by generative AI, with a particular emphasis on the music industry. We have highlighted the pressing need for computational copyright solutions to manage the economic implications of AI-generated content, which could be crucial in tackling regulatory challenges.

Our case studies of existing digital music platforms have set the stage for discussing potential royalty models applicable to AI music generation platforms. Along the way, we have addressed the challenge of attributing the generated music to the copyrighted training content by leveraging data attribution techniques. This study offers a promising prototype of a computational copyright solution in the field of generative AI.

[7]See Section 106 of the Copyright Act: `https://www.copyright.gov/title17/92chap1.html#106`.

[8]See `https://www.copyright.gov/dmca/`.

REFERENCES

21 for 2021: Copyright, re-use and digital business models. https://www.create.ac.uk/blog/2021/06/11/21-for-2021-copyright-re-use-and-digital-business-models/, 2021. Accessed: 2024-01-31.

New york times co. v. microsoft corp et al. U.S. District Court, Southern District of New York, 2023. No. 23-11195.

Timothy Chu, Zhao Song, and Chiwun Yang. How to protect copyright data in optimization of large language models? *arXiv preprint arXiv:2308.12247*, 2023.

Martin Clancy. *Reflections on the financial and ethical implications of music generated by artificial intelligence*. PhD thesis, PhD Thesis. Trinity College, Dublin, 2021.

Hélio Magalhães de Oliveira and RC de Oliveira. Understanding midi: A painless tutorial on midi format. *arXiv preprint arXiv:1705.05322*, 2017.

Eric Drott. Copyright, compensation, and commons in the music ai industry. *Creative Industries Journal*, 14(2):190–207, 2021.

Niva Elkin-Koren, Uri Hacohen, Roi Livni, and Shay Moran. Can copyright be reduced to privacy? *arXiv preprint arXiv:2305.14822*, 2023.

Giorgio Franceschelli and Mirco Musolesi. Copyright in generative deep learning. *Data & Policy*, 4:e17, 2022.

Joshua S Gans. Remix rights and negotiations over the use of copy-protected works. *International Journal of Industrial Organization*, 41:76–83, 2015.

Joshua S. Gans. Copyright policy options for generative artificial intelligence. 2024.

Amirata Ghorbani and James Zou. Data shapley: Equitable valuation of data for machine learning. In *International conference on machine learning*, pp. 2242–2251. PMLR, 2019.

Zayd Hammoudeh and Daniel Lowd. Training data influence analysis and estimation: A survey. *arXiv preprint arXiv:2212.04612*, 2022.

Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the MAESTRO dataset. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1lYRjC9F7.

Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A Lemley, and Percy Liang. Foundation models and fair use. *arXiv preprint arXiv:2303.15715*, 2023.

Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer, 2018.

Andrew Ilyas, Sung Min Park, Logan Engstrom, Guillaume Leclerc, and Aleksander Madry. Datamodels: Predicting predictions from training data, 2022.

Matthew Johnston. How spotify makes money. *Investopedia*, 2023. URL https://www.investopedia.com/articles/investing/120314/spotify-makes-internet-music-make-money.asp. Accessed: 2023-11-27.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *International conference on machine learning*, pp. 1885–1894. PMLR, 2017.

Katherine Lee, A Feder Cooper, and James Grimmelmann. Talkin"bout ai generation: Copyright and the generative-ai supply chain. *arXiv preprint arXiv:2309.08133*, 2023.

Chenghao Li, Dake Chen, Yuke Zhang, and Peter A Beerel. Mitigate replication and copying in diffusion models with generalized caption and dual fusion enhancement. *arXiv preprint arXiv:2309.07254*, 2023.

Daryl Lim. Generative ai and copyright: principles, priorities and practicalities, 2023.

Rachel Manzelli, Vijay Thakkar, Ali Siahkamari, and Brian Kulis. An end to end model for automatic music generation: Combining deep raw and symbolic audio networks. In *Proceedings of the musical metacreation workshop at 9th international conference on computational creativity, Salamanca, Spain*, 2018.

Lee Marshall. 'let's keep music special. f—spotify': on-demand streaming and the controversy over artist royalties. *Creative Industries Journal*, 8(2):177–189, 2015.

Alex McFarland. 9 best ai music generators (december 2023). `https://www.unite.ai/best-ai-music-generators/`, 2023. Accessed: 2023-12-02.

Patrick McKay. Youtube copyfraud and abuse of the content id system. `http://fairusetube.org/youtube-copyfraud`, 2011. Accessed: 2023-11-27.

Elisa Nguyen, Minjoon Seo, and Seong Joon Oh. A bayesian perspective on training data attribution. *arXiv preprint arXiv:2305.19765*, 2023.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

Sung Min Park, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, and Aleksander Madry. Trak: Attributing model behavior at scale, 2023.

Christian Peukert and Margaritha Windisch. The economics of copyright in the digital age. 2023.

Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. Estimating training data influence by tracing gradient descent. *Advances in Neural Information Processing Systems*, 33: 19920–19930, 2020.

Parsa Saadatpanah, Ali Shafahi, and Tom Goldstein. Adversarial attacks on copyright detection systems. In *International Conference on Machine Learning*, pp. 8307–8315. PMLR, 2020.

Matthew Sag. The new legal landscape for text mining and machine learning. *J. Copyright Soc'y USA*, 66:291, 2018.

Matthew Sag. Copyright safety for generative ai. *Forthcoming in the Houston Law Review*, 2023.

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

Malcolm Slaney, Kilian Weinberger, and William White. Learning a metric for music similarity. In *International Symposium on Music Information Retrieval (ISMIR)*, volume 148, 2008.

Anders Søgaard et al. Revisiting methods for finding influential examples. *arXiv preprint arXiv:2111.04683*, 2021.

Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6048–6058, 2023a.

Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Understanding and mitigating copying in diffusion models. *arXiv preprint arXiv:2305.20086*, 2023b.

Katharine Trendacosta. Unfiltered: How youtube's content id discourages fair use and dictates what we see online. *Electronic Frontier Foundation*, 2020.

D Bondy Valdovinos Kaye, Aleesha Rodriguez, Katrin Langton, and Patrik Wikstrom. You made this? i made this: Practices of authorship and (mis) attribution on tiktok. *International Journal of Communication*, 15:3195–3215, 2021.

Ernesto Van der Sar. Youtube processes 4 million content id claims per day, transparency report reveals. https://torrentfreak.com/youtube-processes-4-million-content-id-claims-per-day-transparency-report-reveals- 2021. Accessed: 2023-11-27.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Nikhil Vyas, Sham Kakade, and Boaz Barak. Provable copyright protection for generative models. *arXiv preprint arXiv:2302.10870*, 2023.

A A PRIMER ON THE CONCEPTS OF MUSIC ROYALTIES

It is essential to familiarize ourselves with the fundamental concepts and a few major types of music royalties that are prevalent in the industry.

Streaming Royalty. Streaming royalty arises when music is played on platforms like Spotify, Apple Music, or Amazon Music. Here, copyright owners earn from the platform's revenue, which could be from subscriptions or advertisements. This concept extends to video platforms like YouTube. When copyrighted music is used in the videos, the copyright owners can monetize those videos through advertisements. Streaming royalty generates recurring revenue shares from the platforms.

Synchronization License. Synchronization license usually applies to the use of music in visual media like movies, TV shows, advertisements, video games, and other types of videos (including YouTube videos). The copyright owner grants a license for the music to be synchronized with specific visual content, usually for a one-time payment.

Mechanical Royalty. Mechanical royalties are earned when a song is reproduced[9], such as when a song is downloaded. Mechanical royalties also apply to derivatives of the music. For example, when a DJ or producer creates a remix, they need to obtain mechanical licenses for the songs they are using. Every time the remix is sold, streamed, or otherwise distributed, mechanical royalties are due to the original copyright owners of the songs used in the remix.

It is worth noting that these royalty categories, although insightful for understanding historical practices and industry folklore, do not constitute formal, legal, or mutually exclusive classifications. Furthermore, the outlined descriptions of music royalties are presented at a high level and are considerably simplified. The implementation of royalty models in a real-world context is complex and varies significantly across different platforms. The process of calculating and distributing these royalties involves multiple stakeholders, such as music creators, record labels, and streaming platforms, each with their own agreements and interests. This complexity extends to how revenue is generated——whether through advertisements, subscriptions, or a combination of both.

B DETAILED STAKEHOLDER DESCRIPTION

1. **Artists and Creators:** Musicians, songwriters, and producers who create the content streamed on Spotify.
2. **Record Labels and Music Publishers:** Organizations that own the copyrights to music recordings and compositions.
3. **Music Rights Societies and Collecting Agencies:** Organizations responsible for collecting royalties and distributing them to copyrights owners.
4. **Listeners and Subscribers:** The end-users whose subscription fees and advertising views generate revenue.
5. **Advertisers:** Companies that pay Spotify to advertise on its free-tier platform.

C AI MUSIC GENERATION PLATFORMS BUSINESS MODELS

Services. The backbone of AI music generation platforms is generative AI trained on a large corpus of existing music, which often includes copyrighted music. With the generative AI, the platforms offer a variety of services to meet different needs of end users.

1. **Music Creation and Composition Tools:** These platforms provide advanced AI algorithms that assist users in creating and composing music. This service appeals to a wide range of users, from amateur musicians to professional composers.
2. **Music Libraries and Streaming:** Many platforms host extensive libraries of AI-generated music. These libraries are available for streaming, offering a diverse range of genres and styles to cater to different listener preferences.

[9]This category also includes the case when a song is reproduced in physical forms, such as piano rolls (where the term "mechanical" originates), phonograph records, CDs or vinyl records. But these forms are irrelevant to the purpose of this paper.

3. **Custom Music Solutions:** For corporate clients with specific requirements, such as filmmakers or game developers, these platforms offer custom music generation services. This involves creating unique compositions tailored to the client's needs.
4. **Educational Tools:** Some platforms also include educational components, offering tutorials and resources for learning music composition and production, often incorporating AI tools for a more interactive learning experience.

Stakeholders. The potential stakeholders involved in AI music generation platforms have significant overlaps with those on traditional music platforms, as summarized in the five groups in Section 2.1. The first three groups, including artists and record labels, should be considered for revenue sharing on these platforms. Similarly, end users and advertisers, corresponding to groups 4 and 5, contribute to the platforms' revenue. Additionally, there might be corporate clients paying for custom music solutions.

Revenue Sources. The platforms have several different ways for generating revenues.

1. **Subscription Fees:** Charging users for access to music creation tools, libraries, or streaming services.
2. **Licensing:** Licensing AI-generated music for downloads and other downstream usage. For example, the music may be used in a YouTube video or a video game.
3. **Advertisements:** If the platform offers a streaming service, advertising can be a significant source of revenue.
4. **Custom Composition Fees:** Earning revenue from bespoke music creation services provided to corporate clients or individuals.

D More Details of Experimental Setup

Dataset. We use the MIDI and Audio Edited for Synchronous TRacks and Organization (MAESTRO) dataset (v2.0.0) (Hawthorne et al., 2019) for our experiments. The dataset contains more than 200 hours of piano performances and is stored in MIDI format. The MIDI format is a widely recognized industry standard for representing musical information. Generally, it encodes music as a sequence of MIDI messages. Each type of MIDI message carries different additional attributes. For example, "Note ON" MIDI messages include attributes like velocity, duration, and pitch key. We refer readers to a tutorial (de Oliveira & de Oliveira, 2017) for more details about the MIDI format.

Following the setting of Music Transformer (Huang et al., 2018), we define a vocabulary set of size 388, which includes "NOTE ON" and "NOTE OFF" events for 128 different pitches (256 events in total), 100 "TIME SHIFT" events for different duration lengths, and 32 "VELOCITY" events that set the velocity of the next "NOTE ON" event. The raw data is preprocessed as sequences of events from this vocabulary set.

Model Training and Generation. We train a Music Transformer model on the MAESTRO dataset using the official training set in the dataset. During the training, the music sequences are cropped into a fixed length of $p = 256$ and the model is trained by maximizing the log-likelihood of these sequences. For generation, we use the official test samples in the dataset as prompt to generate music with length $p = 256$. The generated music, denoted as $\{\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_k\}$, will be used to evaluate the data attribution methods. Here $k = 178$ is the number of test samples in the dataset.

Attribution Scores. For the implementation of attribution score calculation, we utilize the TRAK Park et al. (2023) toolkit for both event-level and segment-level attribution. We use the default values in TRAK toolkit for all the hyperparameters. For the ensemble setting, we use 50 models that are independently trained on 50% random subset of the training set.

E Extended Related Work

E.1 Ownership of AI Generation

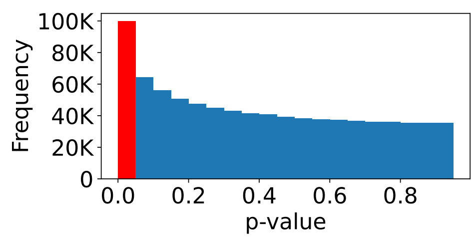
The applicability of fair use influences the determination of ownership and authorship regarding AI-generated content, an issue that remains actively debated and unresolved. The difficulty is to find a human authorship (Peukert & Windisch, 2023). Franceschelli & Musolesi (2022) argues that the author of AI-generated works should be assigned to users. Lee et al. (2023) argues that the generation that does not meet the fair use doctrine should be uncopyrightable[10] unless further legal permission is granted for derivative copyright. This paper adopts the position that the fair use doctrine may not extend to AI-generated content, thereby recommending that authors of training data should engage in negotiations for derivative work permissions.

F ROBUSTNESS OF ATTRIBUTION SCORES

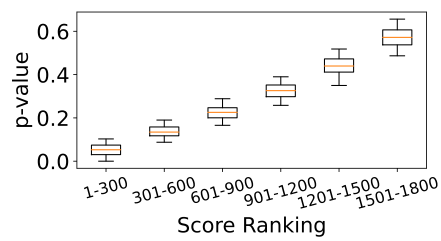
In this section, we analyze the robustness of the TRAK attribution scores for music generative AI, which is crucial for establishing reliable royalty distribution. In Section F.1, we examine the robustness against the randomness inherently existing in the data attribution process, which we term as *stochastic robustness*. In Section F.2, we further investigate the *adversarial robustness* of the attribution scores against malicious actors that seek to adversarially manipulate the attribution scores.

F.1 STOCHASTIC ROBUSTNESS

Data attribution methods for deep neural networks naturally come with randomness due to, e.g., model initialization and training dynamics Søgaard et al. (2021); Nguyen et al. (2023). In this subsection, we examine the stability of the attribution scores against such natural randomness. Specifically, we run the data attribution method multiple times with independent model initialization and training processes. Then we carry out *Student's t-test* on the score of each training-generation pair (with the null hypothesis that the score equals to zero) as a way to quantify the stochastic robustness of the attribution scores.



(a) Histogram of the p-values for t-tests on attribution scores.



(b) P-value v.s. the ranking group of attribution scores.

Figure 2: P-values for t-tests on attribution scores.

Figure 2a shows the histogram of the p-values for t-tests on the training-generation pairs. We find that only a small portion of the pairs have p-values smaller than 0.05[11]. Furthermore, we group the training-generation pairs by the relative rankings of the average attribution scores for these pairs, and Figure 2b shows the boxplots of p-values for each group: the p-value is correlated with the rankings of the attribution scores. This result suggests that while the data attribution scores are not always stable, the ones with the top attribution scores tend to be reliable.

This result has an implication on the royalty mechanism design: the revenue of a generated music should be distributed to the training pieces with top attribution scores.

F.2 ADVERSARIAL ROBUSTNESS

As with any system where financial interests are involved, there might be malicious actors who seek to increase their royalty shares by adversarially manipulating their contributed data. In this section, we evaluate the adversarial robustness of the TRAK attribution scores under two potential adversarial attack methods.

Duplication. Duplicating a training sample multiple times is an intuitive way to increase its total attribution. In the experiment, we add multiple duplicate copies of a training sample into the dataset

[10] See `https://www.copyright.gov/title17/92chap1.html#103`.

[11] Rigorously claiming statistical significance requires false-discovery-rate control for the multiple hypothesis testing. But we are only using the distribution of p-values to measure the stability of attribution scores, and do not intend to claim statistical significance for individual scores.

and then recalculate the attribution scores. Table 2 shows the total attribution score of the duplicated training samples over 178 generated music. We find that having duplications of a training sample in fact mostly decreases the total attribution score of this training sample and its duplications, indicating the attribution scores are robust against this type of attack (see Appendix G for explanations).

Copy-and-Replace. Replacing part of one's music with a segment from a highly influential training sample is another viable method to increase the attribution score of the altered music. In our experiment, we copy a segment from a training sample (source) that has top attribution scores and replace a segment of another training sample (target) with it. Table 3 presents how many times the modified target appears in the top-50 attribution scores among 178 generated music. It indicates that the attack can be effective even with a relatively small number of copied events.

The results of adversarial robustness demonstrate that the current best attribution method, TRAK, is robust to certain adversarial attacks. However, some attacks may still be successful if additional information, such as the attribution scores of other samples, is available. Enhancing the adversarial robustness of attribution scores represents a crucial direction for future research[12].

Duplicate Copies	0 (original)	1	2	5	10	100
Sample 1	-4.187	0.410	-0.054	-0.007	0.115	0.014
Sample 2	0.188	-0.308	0.132	-0.004	-0.016	-0.009
Sample 3	5.128	-0.112	-0.058	-0.054	0.005	0.004

Table 2: The total attribution score of the duplicated sample over 178 generated music. Three training music sequences are chosen to be duplicated independently. Experiments with 1, 2, 5, 10, and 100 extra copies are conducted to compare against the baseline scenario of 0 copies, representing the original setting.

Copied Events	0 (original)	1	2	4	8	16	32
Source 0	0	17	25	24	41	23	33
Source 1	0	19	18	22	21	26	26

Table 3: The number of times that modified target appears in the top-50 attribution scores among 178 generated music (the higher the more influential). Two different music sequences with high attribution scores are selected independently as sources. Experiments on 1, 2, 4, 8, 16, and 32 copied events are conducted to compare against the baseline scenario of 0 events, representing the original setting.

G Experiment Settings and Further discussion for Robustness

Duplication. For the experiment of the adversarial attack using duplication, we add multiple duplicate copies of a training sample into the dataset and then recalculate the attribution scores. The experimental setup, including dataset preparation, model training, generation processes, and data attribution methods, remains consistent with the procedures outlined in Appendix D. We select three independent training music sequences with length of 256 events as the target samples for duplication to assess the sensitivity to the target samples. We test varying quantities of duplicated copies—1, 2, 5, 10, and 100—to explore how changes in the number of copies affect the attribution scores.

This phenomenon that any number of duplicated copies would bring the scores to nearly zero can be understood through the lens of data attribution definition, particularly the concept of leave-one-out influence as described in Eq. 2. The minimal influence observed can be attributed to the presence of at least one duplicated sample in the leave-one-out dataset, thereby diminishing its overall impact.

Copy-and-Replace. For the experiment of the adversarial attack using copy-and-replace, we copy a segment from a training sample (source) that has top attribution scores and replace another training sample (target) at the same position with it. The experimental setup, including dataset preparation, model training, generation processes, and data attribution methods, remains consistent with the procedures outlined in Appendix D. We select two independent training music sequences with top attribution scores as the source samples to be copied to assess the sensitivity to the source samples. We test varying quantities of copied events—1, 2, 4, 8, 16, and 32—to explore how changes in the number of copied events affect the attribution scores.

[12]Please refer to Appendix G for detailed experimental setup.