DOI: 10.71568/daga2026.240

Attribution for Musical Generative AI: A New Perspective

Andrew McLeod[1], Peer-Uli May[2], Thomas Rekittke[2], Hanna Lukashevich[1]
[1] *Fraunhofer Institute for Digital Media Technology IDMT, 98693 Ilmenau, Germany,*
Emails: andrew.mcleod@idmt.fraunhofer.de, hanna.lukashevich@idmt.fraunhofer.de
[2] *Allrights AIxchange UG (haftungsbeschränkt), 10997 Berlin, Germany,*
Emails: pu@allrights-aixchange.com, tr@allrights-aixchange.com

Introduction

As musical generative AI systems have become more prevalent in recent years, questions have arisen about data provenance. Given the recent EU AI Act [2], revenue attribution methods—which distribute license revenues back to the original rights holders of the training data—have become an important part of the ecosystem. Given the projected music generative AI revenues—nearly $8 billion in 2030, according to a projection by GEMA and SACEM, two major European music rights organizations [1]—such methods will soon become an essential component of the musical ecosystem.

Generative AI companies seem to be happy to simply license training data in bulk from labels and collective management organisations (CMOs), leaving them to perform the attribution themselves. However, exactly how to fairly perform such attribution is unclear. In this paper, we analyze this problem of Gen AI Attribution, presenting in particular five prerequisites that a solution must fulfill for it to be practically feasible and accepted by both the music industry and generative AI companies. We further describe existing approaches to the problem and their fulfillment of the prerequisites.

We end by presenting Create Weight Attribution, our proposed solution, derived from Music Information Retrieval (MIR) principles, which is fair, transparent, and fulfills all of the described goals and requirements.

Solution Prerequisites

We divide the solution prerequisites into two categories. First, goals: These are features of a solution that we believe will lead to a healthier and more thriving music industry in general, although a solution without them would still be theoretically and technically feasible. Second, requirements: These features are a necessary component to any solution. Proposals not fulfilling a requirement are infeasible at a fundamental level.

Goals

(1) Fair and understandable. Fair is of course difficult to precisely define, but a reasonable attempt would be: remunerating roughly in proportion with the imparted value of each training piece. Understandable refers to the perspective of the musician or the rights holder. As such, black-box technical solutions, though they might be theoretically or mathematically "fair" (for some definition of fair), do not fulfill this criterion due to their lack of transparency and interpretability. As a whole, this goal ensures that musicians do not feel cheated by the generative AI community.

(2) Emphasis on creativity. This goal is more aspirational in terms of the music industry as a whole. Musical development thrives on creativity and novelty, and a solution that simply emphasizes popularity could lead to a stagnation of new musical ideas. Instead, a solution that explicitly emphasizes creativity and novelty encourages the diversity of music, leading to a more interesting musical landscape, and ensuring that musicians who try something new are also rewarded.

Requirements

(3) No model or training access. Some existing approaches (especially academic approaches) require knowledge of—or even control over—the model or the model training process. For generative AI companies, however, such access and control is untenable. Their model design and training process remain valuable trade secrets, highly optimized for output quality and low cost. As such, any proposed solution must not require access to or control over either the model or the training process.

(4) Model-agnostic. Related to point (3), any solution must be agnostic to the specific type of model used. Although the specifics are unknown, generative AI companies almost certainly use a variety of model architectures and combinations thereof. Therefore, any solution designed to work with only a single type of architecture (e.g., feed-forward networks or Generative Adversarial Networks (GANs)) is unsuitable—even if no direct access or control is required.

(5) Separate master and publishing rights. This final requirement is less technical than the first two, relating instead to the music rights landscape. A song's ownership is typically divided into two categories: the master rights (referring to the specific audio recording of that song) and the publishing rights (referring to the composition itself, e.g., the melody and the lyrics). These rights are usually held and managed by distinct entities, and as such, generated revenues must be able to divided between the two categories.

Existing Methods

This section focuses on existing approaches to attribution, and specifically how they relate to the goals and requirements enumerated in the previous section.

Music Industry

The music industry has existing approaches which they use to remunerate artists from other revenue streams, e.g. from streaming revenues. These include Pro Rata, where

song song receives an equal share of the total revenue (or a share proportional to its length), and Market Share, where each song receives a revenue share in proportion to its popularity. While these methods indeed fulfill the Requirements (assuming some similarly rule-based split is made between the two rights holders for (5)), they fail to meet either of our Goals, being an arbitrary split with no particular emphasis on uniqueness or creativity. While the most popular artists would certainly benefit from Market Share attribution, it is difficult to imagine other artists viewing such terms as fair.

Explainable AI

The vast majority of academic proposals (and a few commercial approaches) come more from the side of model explainability, focusing, for example, on gradient tracking to answer the question: "Why did this model produce this output?" These are, however, limited in terms of their application to the problem at hand. While they are certainly fair, they specifically fail for the Requirements (3) and (4)—if not others as well—requiring significant knowledge of, access to, and control over the model and training procedure.

For example, many academic approaches (e.g., [9]) are restricted to classification tasks, where it is unclear how to transfer the methods to generative AI models outputs. Others require full access to the model and training data in question, or are prohibitively expensive to implement for large models. For example, ensemble strategies (e.g., [7, 8]) propose the training of multiple versions of the original model, each on a carefully-constructed subset of the training data, and using differences between the outputs of the resulting models to gauge the influence of each training data point. Similarly, implicit-differentiation and unrolling-based methods (e.g., [5]) require the computation of specific model gradients for the training points in question (and thus, access to the underlying model). Other methods (e.g., [10]) rely on specific model architectures or training objectives to optimize for attribution accuracy rather than solely output quality.

Output Similarity

Other approaches concentrate less on model explainability per se, but rather attempt to solve the practical question posed to many AI companies currently: "How much money should I pay each training data point?" A common method is to measure output similarity, i.e., the similarity between the generated outputs and each song in the training set. This approach has seen both commercial [4] and academic proposals [6]. Practically, this typically involves training a model to calculate an embedding for each audio file in the training set. Then, for each output generated by a gen AI model, its embedding is calculated and compared to those from the training data, with more similar vectors being assigned a larger proportion of the money attributed to that specific generated output.

Output similarity fares better in regard to our prerequisites. It mostly fulfills Goal (1), being at least more fair, although the embedding models are typically large black-box models, lacking some amount of transparency. It also performs better in relation to Goal (2), where more unusual and creative outputs should generate revenue for the more unusual and creative training songs. For the Requirements, both (3) and (4) are clearly fulfilled, while (5) could be fulfilled if different embeddings are calculated for the musical aspects related to the composer or the performer.

However, another problem remains, namely in the assigning of revenues to the generated outputs directly, which this method necessitates. First, a more philosophical argument: Such an assignment represents the stance that value is imparted from the training data *only at such point in time that the resulting model is used to generate output.* On the other hand, we believe that value is already imparted at train time, and that revenue attribution should be assignable at that point. Second, an analysis of model outputs is not necessarily always possible nor relevant: The outputs may not always be stored or available for measurement, and even if they are, the revenues or license fees due to the rights holders may not be assignable directly to those outputs, e.g. when the generative AI company makes money from user subscription fees rather than the outputs themselves.

Creative Weight Attribution

We instead propose Creative Weight Attribution [3], composed of three pillars: Training Weight, Output Similarity, and Prompt Reference, which can be weighted as desired by rights organisations when calculating their artists' remuneration. Output Similarity, as described above, is not applicable in all cases, but remains nonetheless an informative signal when available. Similarly, Prompt Reference analyses the prompts used to generate outputs (when available), detecting any explicit references to specific artists, genres, or other musical features, adjusting the remuneration proportionally. Training Weight, however, is unique, analysing and remunerating based only on the training data themselves.

In Training Weight, different musical components (e.g., rhythm, bass, harmony, instrumentation, etc.) are measured for each song in the training set. Specifically, Training Weight incorporates over 30 well-founded musicological *Facts*—that is, scalar or vector values which will never change for a given song. Initially, vectors are calculated representing, for example, the rhythm of the bass, or the pitch class distribution of the melody line. From these, a scalar complexity *Fact* is calculated for each representing how much related information a gen AI model could in theory learn from the given song (e.g., a 4-chord loop has potentially less to teach about harmony compared a song with a more complex chord progression).

We also calculate the uniqueness of each vector for the songs in a training set, measuring how much a song has to teach a gen AI model *which that model could only learn from that specific song.* This calculation is performed using kernel density estimation, measuring the local density of the distribution of each type of vector for each song. An example of this calculation is shown in Figure 1 (for hypothetical 2D vectors calculated from a set of 20 hypothetical songs). We then measure the novelty of each

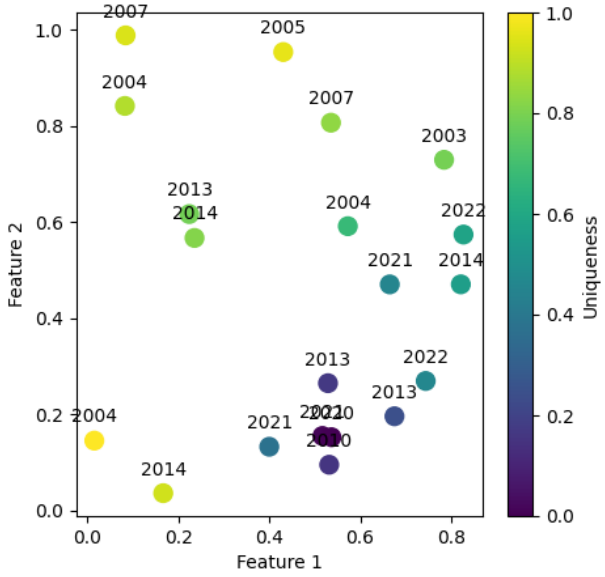

Abbildung 1: Estimate of the uniqueness of data points in a 2D space (representing feature vectors of hypothetical songs) without adjusting by year. Labels represent the release year of each song (which here has no effect).

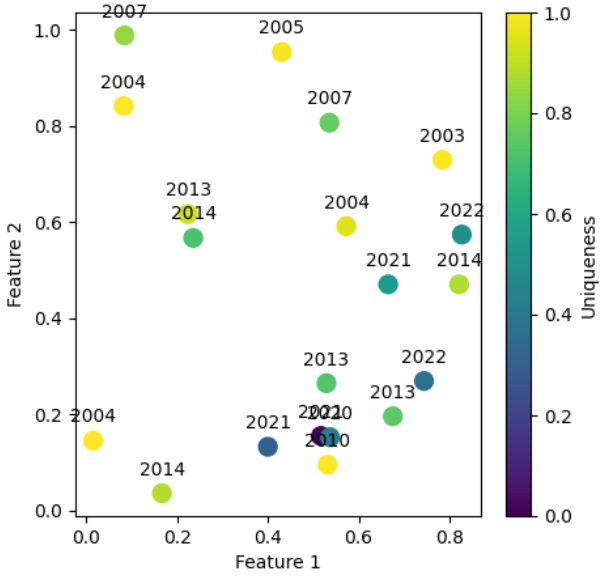

Abbildung 2: Estimate of the uniqueness of data points in a 2D space (representing feature vectors of hypothetical songs), including a year adjustment to reward novelty and creativity. Labels represent the release year of each song.

of these points by adjusting the uniqueness values based on the release year of each song, giving songs which were unique *in their year of release* a higher novelty value. This novelty-adjustment is shown in Figure 2, where, for example, the point labeled 2014 on the right side is now asigned a higher uniqueness value, since it was unique at its release in 2014.

These complexity, uniqueness, and novelty measures are combined with meta-data analysis and measures of human resonance (popularity) of each training track to complete the Training Weight calculation.

Thus, in addition to being applicable to cases where no outputs are available, our method fulfills all of the prerequisites laid out above: It is fair and understandable, being based on established musicological calculations rather than black-box models; the novelty-adjustment ensures an emphasis on musical creativity; it requires no model or training access; it is agnostic to the specific type of model used for generation; and it can be separated between master and publishing rights, where each *Fact* is assignable to one or the other.

Conclusion

In this article, we have described the emerging problem of the fair remuneration of rights holders whose data is used for musical generative AI training. We first listed five prerequisites that we believe a solution to the problem must fulfill, divided between Goals and Requirements. Using these, we then presented existing and proposed solutions to their problem, showing where each fails to fulfill one or more of the prerequisites. Finally, we described our proposed solution, Creative Weight Attribution, which we believe will lead to the fairest, most scientifically-grounded remuneration, and therefore maintain a thriving musical landscape, even as generative AI becomes a more important part thereof.

Literatur

[1] AI and music: Market development of AI in the music sector and impact on music authors and creators in Germany and France. `https://www.goldmedia.com/fileadmin/goldmedia/Studie/2023/GEMA-SACEM_AI-and-Music/AI_and_Music_GEMA_SACEM_Goldmedia.pdf`, 2024. [Online; accessed 31-March-2026].

[2] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), Jul 2024.

[3] Allrights AIxchange UG (haftungsbeschränkt). `https://allrights-aixchange.com/`, 2026. [Online; accessed 31-March-2026].

[4] Sureel: Realtime attribution reporting. `https://www.sureel.ai/services/realtime-attribution-reporting`, 2026. [Online; accessed 31-March-2026].

[5] Juhan Bae, Wu Lin, Jonathan Lorraine, and Roger Grosse. Training data attribution via approximate unrolling. In *Advances in Neural Information*

Processing Systems, volume 37, pages 66647–66686, 2024.

[6] Julia Barnett, Hugo Flores Garcia, and Bryan Pardo. Exploring musical roots: Applying audio embeddings to empower influence attribution for a generative music model. In *ISMIR*, 2024.

[7] Zheng Dai and David K Gifford. Training data attribution for diffusion models. *arXiv Preprint: 2306.02174*, 6 2023.

[8] Junwei Deng, Ting-Wei Li, Shichang Zhang, and Jiaqi Ma. Efficient ensembles improve training data attribution. *arXiv Preprint: 2405.17293*, 5 2024.

[9] Elisa Nguyen, Minjoon Seo, and Seong Joon Oh. A bayesian approach to analysing training data attribution in deep learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 64155–64180, 2023.

[10] Weiwei Sun, Haokun Liu, Nikhil Kandpal, Colin Raffel, and Yiming Yang. Enhancing training data attribution with representational optimization. *arXiv Preprint: 2505.18513v1*, 5 2025.