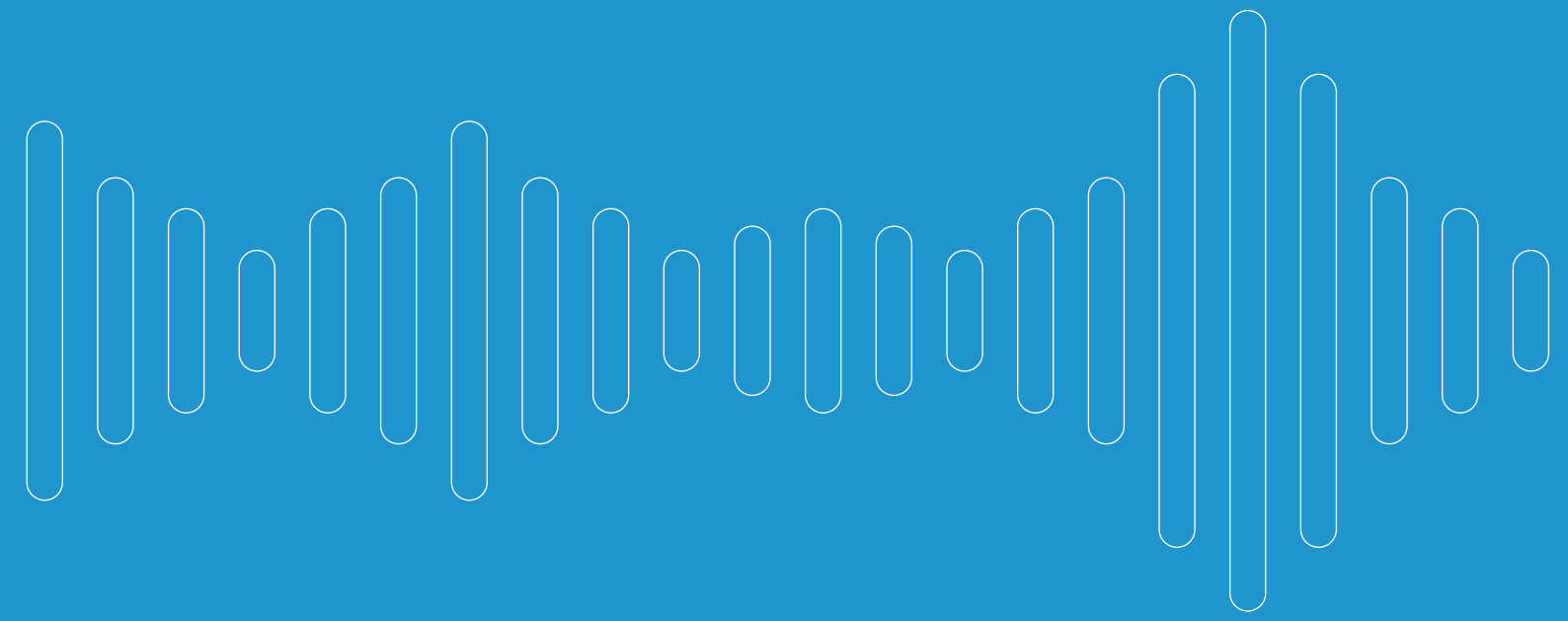
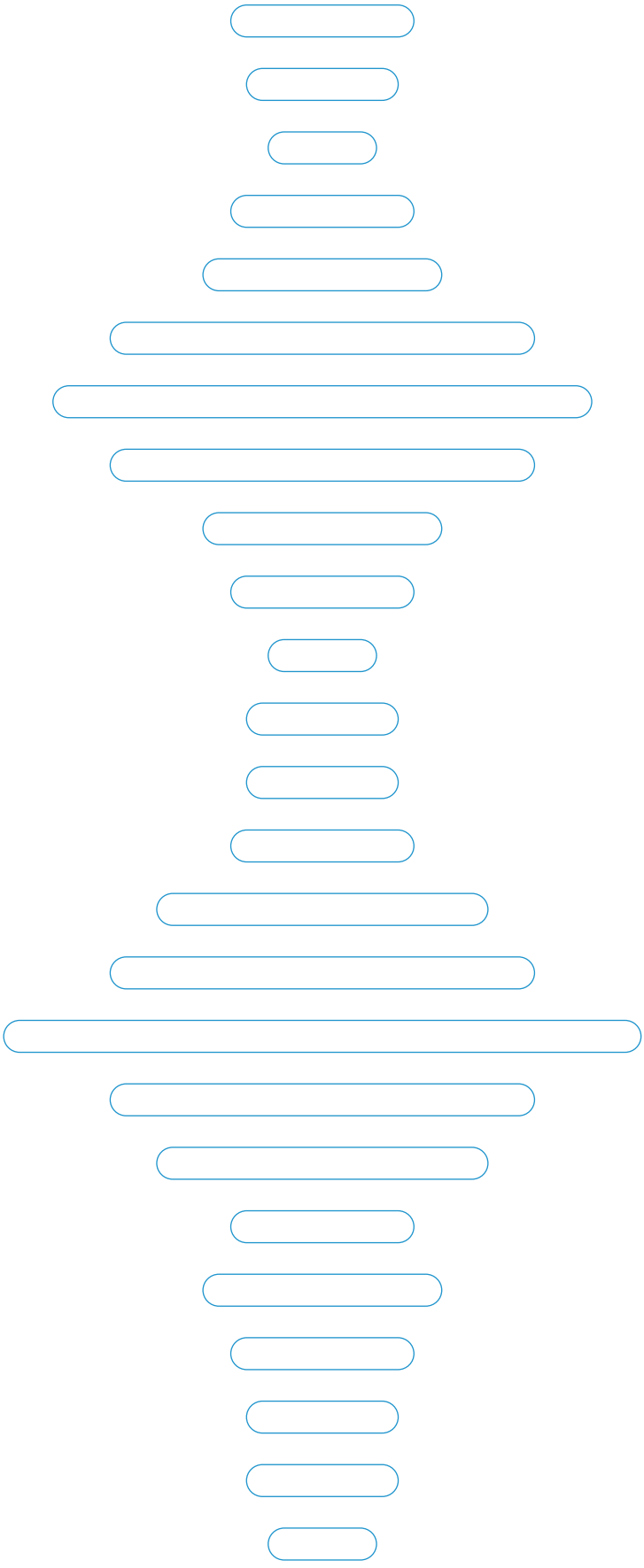
GenAIe

Transparency and attribution in musical AI: Limits and technical possibilities in generative audio models

April 2026



GenAIe
Powered by ale



To cite this publication, use the following bibliographic reference:
AIE and GenAIe (2026): Transparency and attribution in musical AI: Limits and technical possibilities in generative audio models.

Table of Contents

Executive summary

When a generative artificial intelligence model trains with protected works and recordings, the main obligation should be placed on those who design, train and market these systems: documenting training sources, respecting reserved rights, obtaining the necessary authorizations, and activating verifiable remuneration mechanisms.

Based on this premise, this white paper deals with the question that will inevitably follow any training remuneration system, which is how this remuneration can be distributed among the works and recordings that took part in the training.

The problem is not only technical. The opacity of the models solidifies an asymmetry between rightsholders and developers as to the value generated and captured. Protected human catalogues have been used to train generative audio systems without there currently being sufficient mechanisms for transparency, identification, authorisation, attribution and remuneration. Thus, the question about what is inside the models and what the training process has been like are not merely technical questions, but rather a necessary condition for building a balanced and sustainable market for licenses and remuneration.

From a technical perspective, establishing a remuneration system requires usage (which catalogues have been incorporated into the training) and attribution metrics (what influence have these catalogues had on the model's behaviour). This white paper analyses the technologies that exist today or are in development to answer these two questions and evaluate their practical feasibility facing the opacity by design of current models.

We start from the basis that as of today there is no technology ready that is autonomous, deterministic and scalable to identify which works or recordings were used to train a model, what their influence has been, and to make a proportionate distribution on a scale of millions of recordings. The most promising techniques (membership inference, influence functions, machine unlearning) are prototypes that, in addition, generally require access to the internal parameters of the model.

The lack of traceability should not be seen as a natural impossibility, but rather as the result of architectures that have not been designed to be auditable. Without cooperation from the developers, the options are very limited.

European regulation already requires providers of general purpose models to publish information about their training data (art. 53) [14], although the level of detail required today, a structured summary and not a work-by-work inventory, is insufficient. Rightsholders and the collectives related to creators are pushing to strengthen these obligations.

The regulatory priority should move toward standardised, auditable and sufficiently granular registries of datasets, catalogues, and sources used as well as towards effective mechanisms for respecting reserved rights in terms of text and data mining.

In parallel, various agents such as record companies and management organisations are exploring complementary strategies that combine technical pilots for attribution, or litigation, facing model developers and activation of rights on text and data mining. Nevertheless, not any of these as yet has an attribution system for training in production.

In this context, GenAIe, the AIE's Analysis Unit on AI and Music in the Digital Age, has formulated a series of strategic recommendations constructed on three axes:

1

Promotion of **regulatory measures** to broaden transparency obligations and activate effective mechanisms for the reservation of rights in regard to TDM.

2

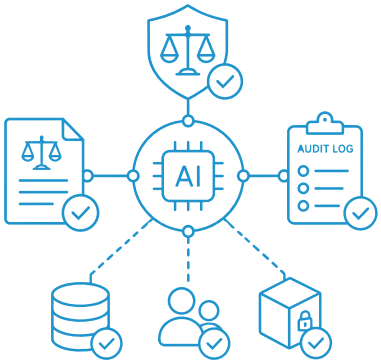
Complementary to this, it is essential to work on quantifying uncertainty through the development of **technical pilots** on indirect attribution based on latent representations (embeddings), probabilistic modelling and detection of anomalies.

These approaches would allow uncertainty to be quantified, to detect aggregated signs of catalogue usage, and to evaluate the relationship between outputs and the catalogue without depending on cooperation from the model developers.

3

Construction of a repertory and audit **infrastructure** that allows taking advantage of any future technical or regulatory advances, in coordination with initiatives and standards at European and international levels.

In the midterm, the strategy should shift from ex post detection toward legal alignment by design: models trained with verifiable registries, effective respect for reserved rights, and auditing abilities incorporated into the architecture. It will be possible, by combining regulation, technical infrastructure, and collective management, to transform a general obligation for remuneration into a distribution system that is proportional, verifiable and defensible for the creators.



THE PROBLEM: WHAT'S INSIDE THE MODEL?

Previous analyses published by GenAIe have dealt with issues such as the **traceability** of AI-generated music content (how to track a file) and **detection** of synthetic content (how to know if something was created by a machine). This white paper addresses a different issue: **Which musical works fed a model's training and how much did each one influence its behaviour?**

Detecting that a song was generated by AI says nothing about which human recordings contributed to the model being able to generate it. A watermark identifies the origin of a specific file, but does not reveal what was in the training dataset. In addition, the effectiveness of audio watermarking has significant technical limitations: A recent study from Sony AI [12] evaluating the robustness of the leading audio watermarking algorithms concludes that **none survives neural compression** (codecs like Descript Audio Codec).

If the watermarks themselves cannot withstand regular transformations in the digital audio pipeline, they are even less useful as a basis for tracking which works fed a training. Responding to this requires technologies that are different from the ones we have analysed to date.

> ***"Detecting that a song was generated by AI says nothing about which human recordings contributed to the model being able to generate it".***

Why it is technically difficult

Current generative AI architectures do not keep a record of what they have learned. During training, the model processes millions of examples and adjusts its internal parameters (hundreds of millions or billions of numbers) to capture statistical patterns from the audio. Once trained, there is no index that we can consult that says "This song by this artist influenced this parameter". Information about training data is distributed opaquely throughout the entire neural network.

The technical report commissioned by the JURI committee of the European Parliament [13] formulates it like this: **current generative AI architectures do not incorporate traceability by design**. What is inside the model is not recorded like an accounting ledger but rather dissolved in a mathematical structure that was not intended to be auditable.

Inferring information about training data is possible but requires indirect, computationally expensive techniques and, in most cases, privileged access to the model.

These limitations explain why solutions require indirect inference techniques, which are detailed in the following section.

THREE TECHNOLOGICAL FAMILIES

Academic and industrial research has produced three main approaches to finding out what's inside a generative AI model. Each responds to a different variant of the question, with different maturity levels and requirements.

Before analysing them, it is important to understand why music raises specific difficulties that do not exist (or are lesser) in text and images.

Why music is a harder problem than text

Most of the research on membership inference and data attribution has been developed for text and image models. Transferring these techniques to music presents additional difficulties:

- **Continuous representation with high dimensionality.** A musical recording is a continuous audio signal, not a discrete token sequence like text. Audio models work with compressed latent representations or spectrograms that encode frequency, amplitude, and phase over time. This representation is much denser than text: one minute of standard quality audio contains millions of numeric values, versus a few hundred text tokens.
- **Multi-level temporal structure.** Music has a hierarchical structure on multiple scales: the timbre texture at millisecond level, the rhythm and harmony at the bar level, the section structure (verse, chorus) at tens of seconds, and the overall shape of the piece at the minute level. Inference techniques developed for text operate on a single temporal scale (the token sequence), making them unsuitable for capturing influence on multiple levels.
- **High redundancy between works.** In text, two documents on the same topic may use similar vocabulary, but the exact sequence of words is usually unique. In music, two recordings of the same genre can share identical harmonic progressions, standard rhythmic patterns, and common instrument timbres. This structural redundancy makes individual membership signals weaker and more ambiguous: the model can "know" a pattern because it saw it in thousands of recordings, not because it saw one in particular.
- **Multiple versions of the same work.** A composition can exist in dozens of different recordings (live versions, covers, remastered works). If the model was trained with one version and another is consulted, detection techniques may fail because the acoustic signal is different even if the work is "the same". For performer management agencies, which manage rights to specific recordings (not compositions), this distinction is critical.

Estas dificultades no invalidan las técnicas que se describen a continuación, pero sí explican por qué los resultados experimentales en música son menos maduros que en texto, y por qué los márgenes de error reportados en la literatura deben interpretarse con cautela cuando se proyectan a escala de catálogo.

Membership Inference: Was this work in the training?

The **MIA** (Membership Inference Attack) technique attempts to determine whether a specific work formed part of a model's training dataset. It is a question with a binary answer, that is, it gives us a yes or a no.

Let's review **how it works**. The work is presented to the model and its behaviour is analysed. If the model "recognises" the work in a statistically distinguishable way from how it treats works it did not see during training, there is evidence of membership. The techniques vary according to the model's architecture:

- In **diffusion models** (those that generate audio by transforming random noise into coherent signal over multiple steps), the most recent approach is LSA-Probe [1]. This method measures two things: the stability of the reverse diffusion path (how much the reconstruction path deviates from a reference) and the level of disturbance needed to degrade the signal (how much additional noise needs to be injected for the model to stop recognising the work). The intuition is that the model follows more stable trajectories and is more resistant to disturbance with works it "knows" from training. This method requires full access to the model's parameters and gradients (white-box access that manufacturers need to provide).
- In **autoregressive models** (such as transformers that generate audio token by token, the architecture used by Suno and Udio), the probabilities the model assigns to each fragment of the work are analysed [2]. If the probability is unusually high compared to works having similar characteristics that were not in the training, there is evidence of membership. This approach can work in black-box mode (only with access to output probabilities), making it more accessible from the outside.
- For **symbolic music** (scores, MIDI), the TS-RaMIA method [3] analyses structural tokens (bar lines, positions, tempo markers) instead of raw audio. Results in controlled environments are promising, although symbolic music represents a small fraction of the problem for organisations that manage audio recording catalogues.

The problem of scale

A MIA is not perfect: sometimes it says "yes, this work was in the training" when it is not true (false positive). In the best published results, the rate of these errors is around 1% [1.2]. That might seem low, but on catalogue scale the figure skyrockets: If you look at a million works, 1% false positives are about 10,000 works wrongly identified as part of the dataset. Likewise, correct detection rates are low (10-20%), so out of every 100,000 works that actually were in the training, only between 10,000 and 20,000 would be detected. The net result is that the system would produce as many false positives as true positives, making it unfeasible for proportional distribution.

Added to this is a specific problem of musical audio: recordings share a lot of structure with each other (common harmonic progressions, standard rhythmic patterns, instrument timbres). Two different performances of the same jazz standard activate similar signals in the model, making it difficult to distinguish whether the model "recognises" a specific recording or has simply learned the pattern shared by thousands of recordings [2].

Dataset Inference: A more realistic alternative

Instead of asking "Was this song in the training?", you can ask "Was this catalogue used?" Dataset Inference [9] aggregates weak signals from many works in the same catalogue to build a more robust statistical test at collection level. The reasoning is that if a catalogue was used to train the model, the individual membership signals (though weak and noisy) should show a consistent statistical bias throughout the catalogue as compared to a control catalogue that was not used.

For large autoregressive models, a recent study [2] reports that between 300 and 900 samples are needed to achieve statistical significance. This is more viable for large labels or management agencies than for individual artists, and is conceptually closer to how collective management organisations work: they manage catalogues, not isolated works.

> ***«Instead of asking "Was this song in the training?", you can ask "Was this catalogue used?" Dataset inference is conceptually closer to how collective management organisations work.»***

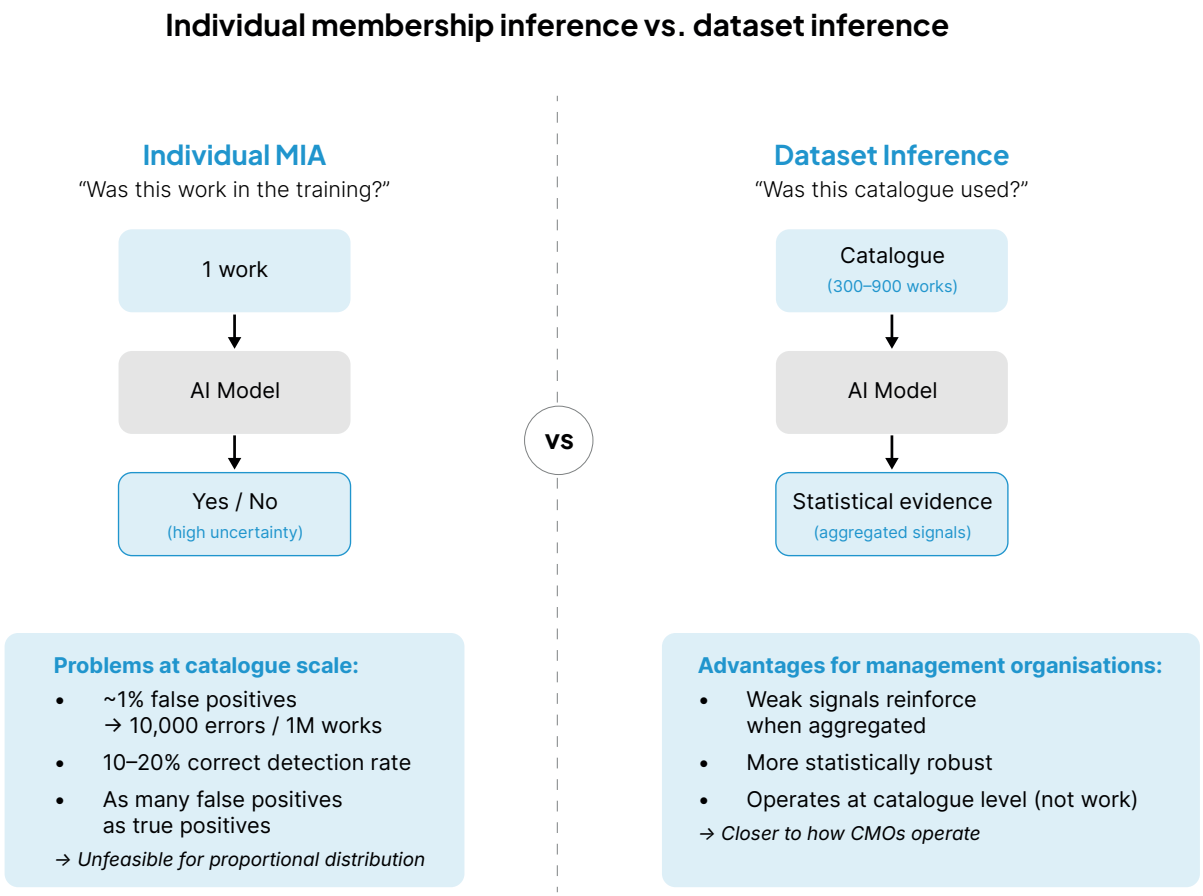


Figure 1 · Comparative diagram MIA vs Dataset Inference. Left: a work → model → yes/no (high uncertainty). Right: catalogue of N works → model → aggregated statistical evidence (greater reliability).

Data attribution and influence: How much did each work contribute?

If the question asked by membership inference is “Was it there?”, **data attribution** asks “How much weight does it have on the result?” Technically, it measures how model behaviour would change if a specific work was removed from the training dataset.

Influence functions

Methods such as Diffusion-TracIn [5] and influence functions based on K-FAC approximations [6] (Kronecker-Factored Approximate Curvature, a technique to efficiently approximate the model’s parameter space curvature) allow us to estimate, for a specific model output, which training examples were most decisive.

The intuitive idea is this: If you generate a song with a certain style and want to know which recordings from the training contributed most to this result, influence functions calculate how the output would have changed if each individual recording had been removed from the training. The recordings whose deletion would have most altered the result are those having the most “influence”. They are the closest technical analogue to “how much each song contributed”.

But this approach is computationally prohibitive. The exact calculation of influence functions requires inverting the model’s Hessian matrix (a matrix of second-order derivatives with as many rows and columns as parameters in the model). For a model with a billion parameters, the matrix would have 10^18 elements. Approximations like K-FAC make the calculation viable, but still the cost is high because the estimate must be repeated for each pair (training work, model output).

The state of the art

In 2025, a Sony AI team published the most directly relevant work for music [4]: they measured the influence of each track on a text-to-music diffusion model trained with 115,000 tracks (~4,356 hours). For each track, they simulated its elimination from training and measured how much the model’s behaviour changed. The system works and produces consistent attribution rankings, but the computational requirements illustrate the scale of the problem: simulating the removal of a single track takes about 20 minutes on an 80GB GPU NVIDIA H100, and going through the entire dataset requires about 5 hours on 8 GPUs H100. For a catalogue of millions of recordings, the cost is multiplied proportionately. A model trained with 10 million tracks (a reasonable order of magnitude for commercial models) would require 90 times more computation, assuming linear scaling.

Sony AI as a research program

It should be noted that this unlearning-based attribution study is not an isolated work, but rather part of a broader Sony AI research program aimed at protecting creator rights in the context of generative AI. In parallel to the work on training data attribution, Sony AI has published CLEWS [11], a system for detecting relationships between versions of musical works at segment level (~20 seconds) that achieves state-of-the-art results in identifying melodic, harmonic and phrasing similarities; and RAW-Bench [12], a robust audio watermarking benchmark whose main finding

is that no current algorithm survives neural compression. Sony presents these three lines as a complementary layered framework [28]: source attribution, detection of relationships between works, and content integrity. This ecosystem approach reinforces a central conclusion of this white paper: There is no technical silver bullet, but rather a set of complementary tools that, combined and with cooperation from developers, could sustain a viable attribution system.

Implementation frameworks

Tools such as TRAK [7] and the dattri library [8] offer infrastructure to operationalise attribution at scale, but are designed for research environments with full access to the model. They do not solve the problem of access to proprietary models such as Suno or Udio.

Implication for management organisations

Influence functions are technically the tool that is closest to what a management organisation needs: quantifying the proportional contribution of each work. But their application requires deep access to the model (weights, gradients, training checkpoints) and a computational cost that only makes sense under a regulated cooperation system. Without the model developer providing access and infrastructure, this family of techniques is inaccessible.

Machine unlearning: the inverse test

A third approach raises the question backwards: If a model can be made to “forget” a work and its behaviour changes measurably, this is evidence that the work was in the training.

Machine unlearning tries to eliminate the influence of specific data without retraining the model from scratch. If by “unlearning” a song the model loses the ability to generate something that it previously produced, the relationship between that song and that output is established.

Important limitations

Recent research warns that unlearning is not always reliable as forensic evidence:

- Models learn structures shared among many works. Removing a nearby work can change the model’s behaviour even if the original work was not in the training (false attribution) [10].
- Different methods of unlearning produce different results, and some create the appearance of unlearning without actually achieving the equivalent effect of retraining without the work [10].
- Specifically in audio, the first systematic analysis of unlearning [10] concludes that the existing methods fail in the most relevant case for management organisations: the removal of individual examples (what the authors call “Item Removal”). The paper, submitted to ICLR 2025 (rejected), has not been validated by peer review, but its main finding is consistent with the general state of the field: unlearning in audio is significantly less mature than in text or images.

Practical value for management organisations

Unlearning can serve as **complementary evidence** in the context of a controlled audit ("by eliminating this work, the probability of generating this motif dropped by X%"), but not as autonomous proof of membership nor as the only basis for a distribution system.

Technology maturity matrix (February 2026)

Technology	What it answers	Maturity	Requires cooperation	Scale (millions)	Validated in audio
MIA – diffusion models (LSA-Probe, 2026)	Was this work in the training?	🔴 Laboratory	**Yes (white-box)** Weights + gradients	🔴 ~1% FPR at scale	Partial (diffusion)
MIA – autoregressive (log-probabilities)	¿Was this work in the training?	🟠 Prototype	**Partial (gray-box)** Output log-probs	🟠 Better with dataset inf.	Yes (Proboszcz, 2026)
Dataset inference (statistical aggregation)	Was this catalogue used?	🟠 Prototype	**Not necessarily** Works black-box	🟠 300–900 samples	Yes (Proboszcz, 2026)
Influence functions (Diffusion-TracIn, K-FAC)	How much did each work contribute to the output?	🔴 Laboratory	**Yes (white-box)** Weights + checkpoints	🔴 115K tracks (Sony)	Yes (Choi, 2025)
Machine unlearning (inverse test)	Does the model change when a work is «forgotten»?	🔴 Laboratory	**Yes (white-box)** Full access	🔴 Not demonstrated	Preliminary (Mason-Williams, 2024)
Creative Weight Attribution (AIxchange / BumaStemra)	What value does each work have for the model?	⚪ Unverified	**No (external)** Metadata + musicology	⚪ Not demonstrated	Not verified

🔴 Laboratory / Not viable at scale 🟠 Prototype / Viable with limitations ⚪ No independent verification

Figure 2 · Table/matrix of maturity of the three technological families. Columns: Technology | What it answers | Maturity (Feb 2026) | Developer cooperation needed | Scale (millions of works). Visual simplification of the main research table.

COOPERATION AS A REQUIREMENT: WHY TECHNOLOGY ALONE IS NOT ENOUGH

One pattern crosses the three technological families: **most effective techniques require some level of access to the model**. Without it, the options for a management organisation are reduced to black-box approaches with insufficient precision for proportional distribution.

The technical briefing from the European Parliament [13] proposes this as the main conclusion: **business policy should not be based exclusively on adversarial reverse engineering**. Inference and attribution techniques function as auditing instruments within a framework of mandatory transparency, not as autonomous tools.

With no cooperation: What a management organisation can do today

If a management organisation only has access to public generation services (generate audio from prompts, without access to internal parameters), its technical options are:

- **Dataset inference** at catalogue level, adding samples to detect if an imprint or collection was used. This requires hundreds of samples per consultation and the results are fragile in the face of changes in distribution.
- **Similarity analysis** between generated outputs and works in the catalogue, using musical embeddings. This measures similarity, not causality, but can generate useful circumstantial evidence.

These tools are useful for occasional disputes and negotiations, but not for a mass distribution system.

With partial cooperation: What opens up

If the companies developing the generation models allowed consultation of certain internal data from the model without giving full access (for example, what probability assigned to each audio fragment, or endpoints designed for auditing):

- Grey-box MIAs would gain significant precision.
- Dataset inference would become more reliable and require fewer samples.
- Standardised sampling protocols could be designed.

With full cooperation: What would be possible

If suppliers provide audited access to weights, gradients and training logs, directly or through a certified third party:

- Influence functions could quantify proportional contribution.
- Unlearning protocols could validate attribution.
- A distribution system based on technical evidence could be built.

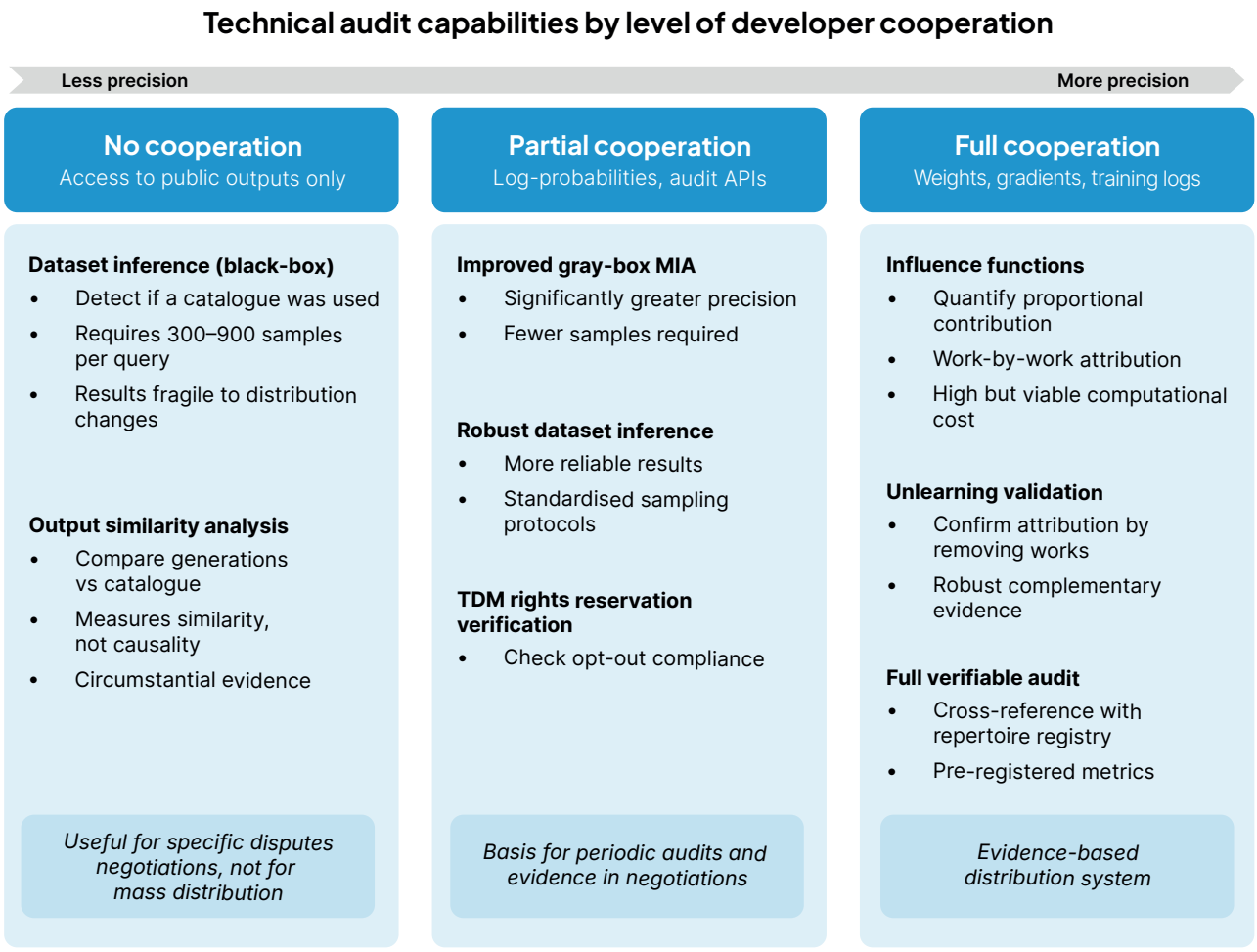


Figure 3 · Three-column framework (no cooperation / partial cooperation / full cooperation) showing what is technically feasible in each scenario. It reinforces the message that regulation is the key enabler.

THE REGULATORY FRAMEWORK: LEVERS TO FORCE COOPERATION

The technical feasibility of attribution depends on cooperation from the developers. What can obligate them or compel this cooperation?

The EU AI Act: What already exists

European regulation on AI (2024/1689) [14] lays down specific transparency obligations for general purpose models (GPAI) in **Article 53**:

- **Art. 53.1.c:** providers of general-purpose AI models shall put in place a policy to comply with Union law on copyright and related rights, and in particular respect the reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 on Copyright.
- **Art. 53.1.d:** providers must draw up and make publicly available a sufficiently detailed summary about the content used for training, according to a template provided by the AI Office.

The template, published 24 July 2025 [15], requires information on content types, sources (datasets, scraping, user data), periods and main domains. It does not require a work-by-work inventory.

Does it apply to music generation services like Suno or Udio? Yes, if they market a GPAI model in the EU. The Regulation itself clarifies (recital 97) that obligations apply even when the provider integrates their own model into their own product.

Timetable for application:

- Obligations in force as of **2 August 2025** (art. 113).
- Effective enforcement: new models as of **August 2026**; existing models up to **August 2027** for adaptation.
- Penalties: up to **15 million euros or 3% of global annual turnover** for non-compliance.

The gap between what is required and what is necessary

The level of detail required by the AI Act today (a structured summary of sources and modalities) is a first step, but **does not meet the needs of a management organisation**. A summary that says "musical audio datasets were used from commercial and public sources" does not allow identification of which specific recordings were used nor calculation of proportional contributions. A report by the Danish Rights Alliance [26] confirms that none of the major AI providers is sufficiently transparent about their training data, and that voluntary transparency is decreasing over time.

The music sector is quite clear on this. A coalition of rightsholders including CISAC, GESAC and IFPI [17] called the July 2025 implementation package (code of practice, guidelines and template) **insufficient** for effective exercising of rights. Their position: without significant transparency about the specific content used, no functional licensing market is possible.

The Copyright Directive as a complement

Copyright Directive 2019 (2019/790) [16] establishes in articles 3 and 4 the conditions under which text and data mining (TDM) is permitted:

- **Article 4** allows commercial TDM provided that the rightsholder has not reserved their rights "appropriately", including machine-readable means in the online environment.
- Organisations such as SACEM and GEMA have already activated these reservations, forcing developers to negotiate licenses if they want to use their repertoire.

The AI Act does not create a new license, but rather requires providers to document how they respect these reservations of rights, which strengthens the negotiating position of management organisations.

Ongoing legislative proposals

In the USA, two proposals are moving forward in Congress:

- The TRAIN Act (July 2025) [29] would create an administrative subpoena process to assist copyright owners in determining which of their copyrighted works have been used in training.
- The CLEAR Act (February 2026) [30] proposes transparency obligations similar to the European ones.

None are yet in force, but they reflect a global trend toward mandatory transparency.

Regulatory framework: timeline

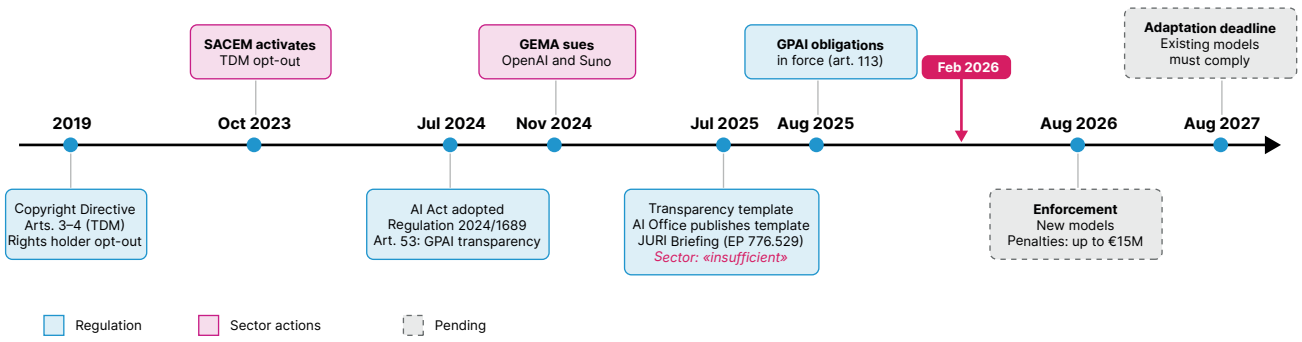


Figure 4 · Horizontal timeline: Copyright Directive (2019) → AI Act adopted (July 2024) → Transparency Template (July 2025) → GPAI obligations enter into force (August 2025) → Enforcement for new models (August 2026) → Enforcement for existing models (August 2027). Highlight the current point in time (February 2026).

WHAT OTHER MANAGEMENT ORGANISATIONS ARE DOING

No management organisation is facing this problem alone. A mapping of the main collective management organisations shows a sector that is advancing on three simultaneous fronts: registration control, legal enforcement, and technical exploration.

Technical pilots

- **BumaStemra** (Netherlands) is the organisation that has publicly made the most progress in technical attribution. Since the beginning of 2025, it has registered works partially generated by AI with a specific identifier [21], and collaborated with the startup AIxchange (Berlin) [19] on a "Creative Weight Attribution" model that aims to measure the contribution of each work to the model. The pilot was reported in industry press in January 2026, although there is no official statement from BumaStemra confirming implementation nor results.
- **PRS for Music** (United Kingdom) has anomaly and fraud detection tools in works registries, although not a training attribution system. It has been promoting the Creative Rights in AI Coalition since December 2024 [31].

Legal enforcement

- **GEMA** (Germany) has opted for judicial recourse: it sued OpenAI in November 2024 [22] and Suno in January 2025 [23]. In parallel, it published an AI Charter with ten principles and activated TDM reservations.
- **SACEM** (France) exercised its TDM opt-out in October 2023 [24], establishing that training models with their repertoire requires prior authorisation.

Sectoral coordination

- **CISAC** (global organisation, 227 member companies) maintains an AI working group that has been active since at least 2020 [32] and ongoing technical committees for standards and distribution.
- **GESAC** (European grouping of author management organisations) leads the regulatory pressure put on European institutions.
- **ASCAP, BMI and SOCAN** (USA and Canada) aligned their works registration policies with AI components in October 2025 [25].
- **SCAPR** (international organisation for collective management of performers' rights) has not

yet pronounced an identified public position on technical pilots for training attribution. This situation highlights a strategic opportunity for the performer ecosystem: **to promote a primary common technical position on attribution applied to related rights, anticipating regulatory developments and contributing to defining international standards in a still emerging field**. In this context, the most advanced positions in the performer ecosystem come from AEPO-ARTIS.

«No collective management organisation has a training attribution system in production. The sector builds legal, technical and institutional infrastructure when technology or regulation allow»

E "CREATIVE WEIGHT ATTRIBUTION" APPROACH

AIxchange (Berlin) has coined the term "Creative Weight Attribution™" (CWA) and has positioned itself as an industry solution for fair distribution of AI training. This should be analysed separately because it is the most specific initiative and the one most closely linked to the management organisation field.

Based on public information available (AIxchange website, an article in Digital Music News from January 2026, and a conference abstract by Andrew McLeod from Fraunhofer IDMT, scheduled for DAGA 2026 [20]), Creative Weight Attribution evaluates each work in the training according to multiple musical characteristics (complexity, uniqueness) combined with metadata analysis and popularity ("human resonance"). McLeod's abstract notes that methods claiming to calculate attribution based on model weights "**have not yet been demonstrated as being possible**", and proposes CWA as a pragmatic alternative. It should be noted that it has not been possible to independently verify the technical details of CWA beyond what AIxchange and its collaborators publish.

CWA, as publicly described, looks more like a **data valuation policy** (how to weigh works for distribution once it is known which ones are in the training) than a **discovery technology** (finding out which works are in the model). It can be useful for allocating payments once inclusion is established, especially in subscription income models where individual output accounting is difficult. But it should not be confused with a demonstrated technical ability to extract which songs trained a model based on its parameters.

Recommendations

On the basis of this analysis, GenAIe formulates three axes for action prioritised by time periods.

Short-term (6–12 months)

1. Regulatory pressure: expand transparency obligations. The AI Act already establishes a framework, but the level of detail required is insufficient for distribution. It is recommended that management organisations, in coordination with CISAC, GESAC and other performer management agencies, promote:

- That the transparency template makes it obligatory to identify **musical audio sources** in a more granular way (specific datasets, licensed catalogues, scraping sources with sufficient detail to cross with repertoire bases).
- That enforcement mechanisms from the AI Office include effective verification of these obligations.

2. Activate TDM rights reservations. Following the example of SACEM and GEMA, it is recommended that management organisations formally exercise the reservation of rights under Article 4.3 of the Copyright Directive for the repertoire they manage, stating that any use of recordings from their artists to train models requires authorisation.

3. Evaluate dataset inference pilots. It is recommended to launch a Dataset Inference technical pilot against the main publicly available music generation models. The goal is not a distribution system, but rather to generate negotiating evidence: to demonstrate that certain catalogues were used in the training.

Mid-term (1–3 years)

4. . Build repertoire and audit infrastructure. To take advantage of any technical or regulatory advances, the sector needs to move forward in:

- A **canonical repository** of recordings linked to standard identifiers (ISRC, IPI, ISWC) that allows crossing with audit results.
- **Secure computing** ability, directly or through consortia, to perform technical audits when access to models is gained.
- **Standardised** audit **protocols** with pre-registered metrics, fixed thresholds and matched control sets.

Long-term (3+ years)

5. Alliances with other management organisations and technical consortia. It is recommended to explore collaboration with initiatives such as BumaStemra/AIxchange to evaluate CWA as a complementary valuation model, as well as with CISAC to advance technical auditing standards. Consider participation in initiatives such as the Data Provenance Initiative [27] for tracking of datasets.

6. Combined evidence-based distribution system. When the regulatory framework and technological maturity permit, it is proposed to move toward a system that combines:

- Mandatory disclosure of training datasets (regulation).
- Technical verification through inference and attribution (audit).
- Weighted contribution assessment (distribution policy, potentially inspired by CWA-style models or on distribution proposals based on musical similarity [18]).

This system will require coordinated institutional, technical and regulatory collaboration at European and international level.

REFERENCES

Academic and technical sources

1. Liu, Y. et al. (2026). “Membership Inference Attack Against Music Diffusion Models via Generative Manifold Perturbation (LSA-Probe).” arXiv. https://arxiv.org/pdf/2602.01645
2. Proboszcz, J. et al. (2026). “Membership and Dataset Inference Attacks on Large Audio Generative Models.” arXiv / NeurIPS 2025 Workshop. https://arxiv.org/pdf/2512.09654
3. Liu, Y. et al. (2026). “TS-RaMIA: Membership Inference Attacks for Symbolic Music Generation Models.” PMLR EAIM@AAAI. https://openreview.net/pdf?id=VeI3Nt6Hck
4. Choi, W. et al. (2025). “Large-Scale Training Data Attribution for Music Generative Models via Unlearning.” Sony AI. https://arxiv.org/html/2506.18312v2
5. Xie, T. et al. (2024). “Data Attribution for Diffusion Models: Timestep-induced Bias in Influence Estimation (Diffusion-TracIn / ReTrac).” TMLR. https://arxiv.org/abs/2401.09031
6. Mlodozeniec, B. et al. (2025). “Influence Functions for Scalable Data Attribution in Diffusion Models.” ICLR 2025 (Oral). https://arxiv.org/abs/2410.13850
7. Park, S. et al. (2023). “TRAK: Attributing Model Behavior at Scale.” ICML. https://github.com/MadryLab/trak
8. Deng, J. et al. (2024). “dattri: A Library for Efficient Data Attribution.” NeurIPS 2024. https://arxiv.org/abs/2410.04555
9. Maini, P. et al. (2024). “LLM Dataset Inference: Did you train on my dataset?” arXiv. https://arxiv.org/abs/2406.06443
10. Mason-Williams, I. et al. (2024). “Machine Unlearning in Audio: Bridging The Modality Gap Via the Prune and Regrow Paradigm.” Presentado a ICLR 2025 (rechazado). https://openreview.net/forum?id=i3tBySZWrR
11. Serrà, J. et al. (2025). “CLEWS: Contrastive Learning from Weakly-Labeled Segments for Musical Version Identification.” ICML 2025. Sony AI. https://arxiv.org/abs/2502.16936
12. Özer, Y. et al. (2025). “RAW-Bench: Benchmarking Robustness of Audio Watermarking.” INTERSPEECH 2025. Sony AI. https://www.isca-archive.org/interspeech_2025/ozer25_interspeech.html

Institutional and regulatory sources

13. European Parliament, JURI (July 2025). "Technological Aspects of Generative AI in the Context of Copyright." PE 776.529. https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/776529/IUST_BRI(2025)776529_EN.pdf

14. 14. Directive (UE) 2024/1689 (AI Act), art. 53, 113, recitals 97, 107, 108. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-53

15. 15. European Commission (July 2025). Template public summary training content for GPAI models. https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models

16. Directive (UE) 2019/790, arts. 3 and 4. https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng

17. IFPI / CISAC / GESAC et al. Joint declaration on implementation measures of the AI Act (2025). https://www.ifpi.org/coalition-of-eu-creators-and-rightsholders-raises-serious-concerns-about-the-third-draft-of-the-general-purpose-ai-code-of-practice-under-the-eu-ai-act-no-code-is-better-than-the-fundamentally-flawed/

18. Herremans, D. (2025). "Royalties in the age of AI: paying artists for AI-generated songs." WIPO Magazine. https://www.wipo.int/en/web/wipo-magazine/articles/royalties-in-the-age-of-ai-paying-artists-for-ai-generated-songs-73739

Industry and management organisation sources

19. AIxchange. "Creative Weight Attribution™." https://allrights-aixchange.com/

20. McLeod, A. / Fraunhofer IDMT (marzo 2026). "Attribution for Musical Generative AI: A New Perspective." Conference DAGA 2026. https://conforg.fr/bin/v2/htmlview?dir=daga2026&lang=1&ref=240

21. BumaStemra – AI works registry. https://bumastemra.nl/ai-geregistreerd-werk/

22. GEMA – Lawsuit against OpenAI (November 2024). https://www.gema.de/de/w/gema-erhebt-klage-gegen-openai

23. GEMA – Lawsuit against Suno (January 2025). https://www.gema.de/de/w/pm-klage-gegen-suno

24. SACEM – TDM opt-out (October 2023). https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out

25. SOCAN / ASCAP / BMI – Alignment on AI registration policies (October 2025). https://www.socan.com/ascap-bmi-and-socan-announce-alignment-on-ai-registration-policies/

26. Danish Rights Alliance (September 2024). "Report on AI model providers' training data transparency." https://rettighedsalliancen.dk/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf

27. Data Provenance Initiative. https://www.dataprovenance.org/

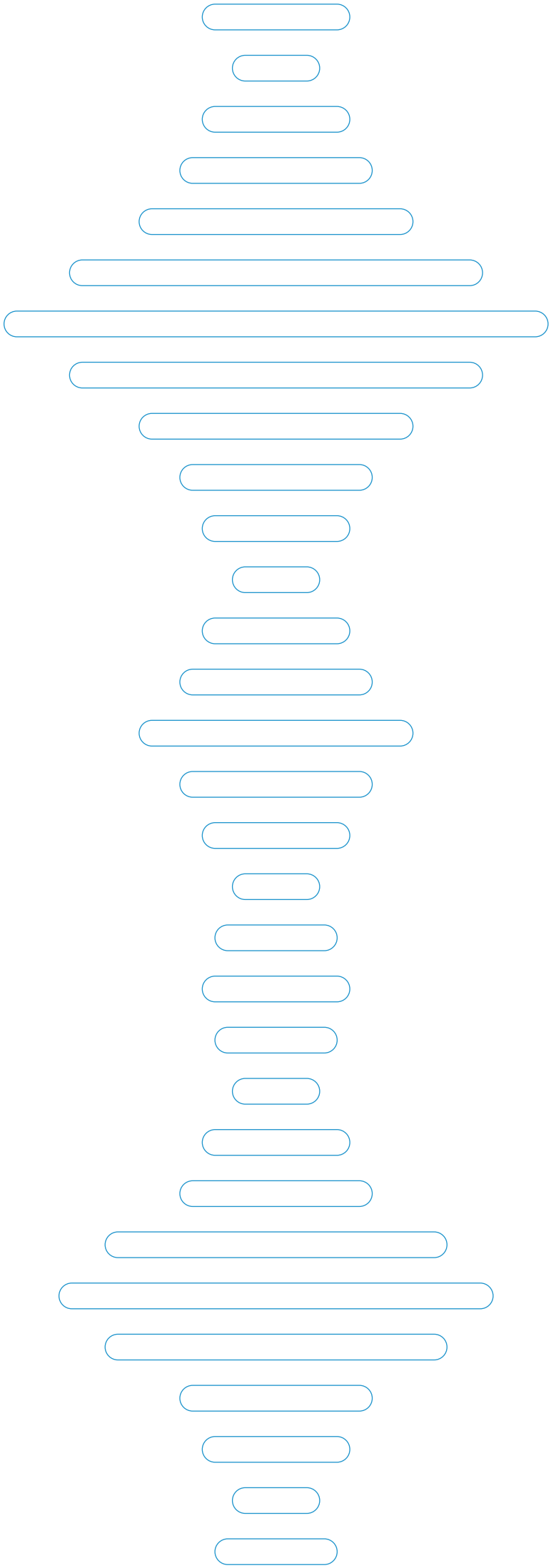
28. Sony AI (2025). "Protecting Creator's Rights in the Age of AI." Blog Sony AI. https://ai.sony/blog/Protecting-Creator%E2%80%99s-Rights-in-the-Age-of-AI/

29. TRAIN Act, S.2455, 119th Congress (July 2025). https://www.congress.gov/bill/119th-congress/senate-bill/2455/text

30. CLEAR Act (February 2026). Sens. Schiff and Curtis. https://www.schiff.senate.gov/news/press-releases/news-sens-schiff-curtis-introduce-bipartisan-bill-to-protect-creators-work-implement-transparency-safeguards-in-ai-model-development/

31. PRS for Music – Creative Rights in AI Coalition (December 2024). https://www.prsformusic.com/press/2024/creative-rights-in-ai-coalition-calls-on-government-to-protect-copyright

32. CISAC – Artificial intelligence policy. https://www.cisac.org/services/policy/artificial-intelligence



GenAIe

Powered by

www.genaie.es