



Music Metadata: The Hidden Infrastructure of Fairness

DIGITAL RIGHTS AWARENESS MODULE 8

Author:

Giuseppe Mazziotti Coordinator and Principal Investigator of the Fair MusE project



Funded by
the European Union

A decorative vertical bar on the left side of the slide, composed of various geometric shapes including circles, triangles, and semi-circles in colors like green, orange, pink, and grey, some with patterns like stripes or dots.

Text and Data Mining: The Starting Point

- AI models learn through **Text and Data Mining (TDM)** — automated analysis of massive data sets to detect patterns.
- In music, this means scanning thousands of recordings to learn styles, rhythms, or production traits.
- The **2019 DSM Directive** created two TDM exceptions — meant for research, not large-scale AI — but now central to this debate.
- When no exception applies, **a licence is required** — raising the question: does AI training count as TDM?

Two Paths for TDM

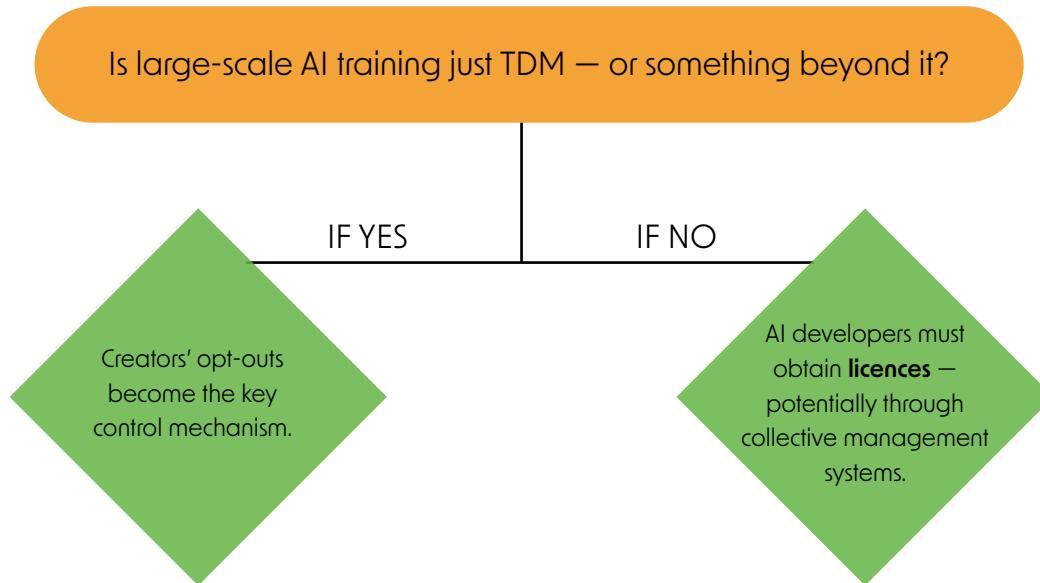
ARTICLE 3	ARTICLE 4
Allows research organisations and cultural institutions to mine data for scientific purposes if they have lawful access.	Allows TDM for any purpose , commercial or not, unless rights are reserved by the copyright holder.

But “rights reservation” rules differ across Europe — some demand **machine-readable opt-outs**, others accept simple notices.

The 2024 **LAION ruling in Hamburg** recognised that natural-language opt-outs can also be valid if detectable by crawlers.

This uncertainty leaves creators unsure whether notices, metadata, or robots.txt files actually prevent scraping.

The AI Training Dilemma



Either way, new systems for remuneration and licensing are needed to keep creators in the loop.



Originality, Authorship, and Derivative Use

Under EU and international law, only **humans can be authors**.

- AI-generated works without meaningful human input are **not protected by copyright**.
- Yet most GenAI outputs are **collaborative**, combining human prompts with machine-generated content.

Are prompts creative acts — or mere instructions? The law is unclear.

- Another question: when does an AI output **infringe** pre-existing works?

The **Pelham case (2019)** held that sampling is lawful only if the reused sound is **not recognisable**.

- Applied to AI, this means the key issue is how close outputs are to the works that trained the model.

Legal and Policy Developments

Cases across the world now test these boundaries:

GEMA v OpenAI and GEMA v Suno
in Germany.

**RIAA and major labels v Suno and
Udio** in the U.S.

Getty Images and New York Times
suing over uncensored scraping.

Was the training lawful?

Were rights reservations respected?

In the U.S., "**fair use**" may shield some training if deemed **transformative** and non-substitutive (as in Authors Guild v. Google, 2015).

But fair use doesn't exist in EU law, so Europe must rely on specific exceptions and licensing frameworks.

The EU's Answer: The AI Act

The AI Act (2024) introduces copyright compliance duties for general-purpose AI.

Article 53(l)(c)

AI providers must adopt a copyright-compliance policy, respecting DSM Directive opt-outs.

Article 53(l)(d)

They must publish summaries of training data.

These obligations apply even if training occurs **outside the EU** — binding companies like OpenAI or Google.

The **EU AI Office** oversees compliance and coordinates a new **Code of Practice** signed by major AI firms.

The aim: shift responsibility toward **transparency and lawful data use**.



Fairness, Licensing, and the Data Gap

Two levers define how AI and music copyright can coexist:

Input phase

Rights reservations and collective licences to authorise training.

Output phase

Content marking and attribution under **Article 50 of the AI Act.**

Both depend on reliable, auditable **data and metadata.**

Without interoperable identifiers, usage logs, and linked repertoires, **fair remuneration remains impossible.**

In recorded music, fairness is inseparable from data **interconnection** — the infrastructure that ensures every creative layer is recognised and rewarded.



Find us on



fairmuse.eu



Creative Commons' licence (BY NC ND 4.0)



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.