



## Digital Rights Awareness

### Episode 8

## AI, Generative AI & Creators' Rights

### Script

Author:

**Giuseppe Mazziotti** - Coordinator and Principal Investigator of the Fair MusE project



This project is funded by the European Union's Horizon Europe Framework Programme programme, under the Grant Agreement no: 101095088

From sheet music to streaming, from ownership to algorithms, our journey through the evolution of music rights has shown how every technological leap reshapes creativity, access, and the logic of fairness.

Now, we stand at the edge of a new frontier: artificial intelligence. Machines can learn from millions of human creations and generate music, images, or words in seconds. In doing so, these technologies challenge the very foundations of copyright and force us to address a complex question.

When creativity becomes computational or is heavily assisted by machines, what remains uniquely human?

In this final chapter of our series, we explore how EU law is responding to this new reality — from text-and-data mining rules to recently adopted provisions in the 2024 AI Act — and how creators can remain visible and protected in an automated world.

## Text and Data Mining: The Starting Point

Before machines can generate anything, they must learn. This learning process occurs through a technical process that implies the ingestion and processing of large amounts of data. As current debates are showing, such preliminary activities performed by machines might qualify as ‘text and data mining’ (TDM). This is the automated computational analysis of large collections of digital information to identify patterns, trends, or relationships. In the context of music, this could mean analysing dozens of thousands of songs or recordings to detect rhythm patterns, chord progressions, or production styles that will later train an AI model.

In 2019, the EU introduced two mandatory copyright exceptions for TDM through the well-known DSM Directive. These copyright exceptions aim to foster research, innovation, market integration, and legal certainty. These rules pre-date the GenAI surge, yet today they sit at the center of disputes about large-scale scraping and ingesting of creative works.

As a general principle in copyright law: where no exception applies, a licence is required for the use of a copyright work to be legitimate.

So, the dilemma is: do TDM exceptions apply to the training of machines based on scraping and ingesting large volumes of copyright works? And, if not, how shall AI developers obtain a licence for the use of these works?

## Two Purposes (and Potential Paths) for Text and Data Mining Research Purposes (Article 3 DSM Directive)

Article 3 of the DSM Directive allows research organisations and cultural heritage institutions to mine data for scientific purposes, provided they have lawful access. Most national transpositions follow this model, though details differ.

Ireland, for instance, stands out. Irish law allows any person with lawful access to perform web crawling for non-commercial research, while imposing acknowledgement and notification duties - a progressive but difficult-to-enforce approach.

Across the EU, at least one principle is clear: *lawful access* preserves right-holders' ability to license catalogues while supporting genuine research.

## All Other Purposes (Article 4 DSM Directive)

Article 4 of the DSM Directive is broader than the previous provision: this provision permits TDM for any purpose, commercial or not, *unless* right-holders explicitly reserve their rights under Article 4(3).

This 'rights reservation' requirement has caused uncertainties at the national level. It is still unclear under the laws of EU Member States what shall count as an effective reservation. Some States require *machine-readable* notices from copyright holders; others are looser or just vague.

In September 2024, the Hamburg Regional Court in Germany ruled in its LAION case that machine-readability is a *dynamic* concept reflecting the state of technological development. Therefore, if crawlers are capable of reliably detecting a natural-language reservation, courts may recognise such notices as sufficient TDM opt-outs.

For composers, performers, and record labels, this uncertainty makes it hard to know whether a simple website notice, metadata, or only a robots.txt file — which is considered the standard option for machine-readable opt-outs — actually outlaws scraping.

## The AI Training Dilemma

If we consider what has been said so far, it is easy to understand that a central question in this domain is whether large-scale AI training qualifies as TDM under Article 4 or falls outside its scope.

## Episode 8 - Script

On the one hand, the answer could be affirmative if we consider ‘TDM’ according to the broad definition provided under the DSM Directive: it’s ‘computational analysis’ and, so, can include model training.

On the other hand, the answer could be negative - or more difficult to reach - if we consider that, back in 2019, EU lawmakers did not contemplate AI and, even less so, today’s general-purpose AI models’ scale and their economic impact.

To sum up, if training counts as TDM, then effective opt-outs by copyright holders become the key tool for control. If it doesn’t, then AI developers must seek ‘blanket’ licences from collective management organisations or owners of large content libraries, with the necessity to develop new systems of remuneration.

## Originality, authorship, derivative use

International and European copyright law still rests on the principle that only humans can be considered ‘authors’ of creative works. A work must be the result of *human intellectual creation* to be protected. Autonomous machine outputs, without meaningful human input, presently fall outside protection.

But most GenAI uses are collaborative in nature and entail different layers of ‘authorship’, some of which are instigated by users’ prompts, selections, and output refinement. Whether this contribution meets the copyright law’s originality threshold is unclear or very difficult to assess.

For instance, are prompts acts of creative expression, or simply functional instructions?

Equally relevant and tricky are questions (and possible legal disputes) arise over the concept of ‘derivative’ use: when does an AI-created output infringe rights in pre-existing works?

This question is even more difficult to address in a legal framework, such as that of the European Union, where the law has not harmonised (yet) the concept of “derivative work” so that national courts are free to determine what makes a new work substantially similar to (or derived from) an earlier work.

At European level, the EU Court of Justice, in the famous 2019 ruling in the case “*Pelham GmbH and others v. Ralf Hütter and Florian Schneider-Esleben*”, concerning the legitimacy of sampling of a sound recording without the authorisation of the record producer, held that such a creative re-use can be free, and therefore lawful, if the inclusion of the earlier work into the new one is *not* recognisable; otherwise, sampling requires a licence from the original recording’s creator.

## Episode 8 - Script

If applied in the AI domain, this would raise the question of whether AI's creative outputs are sufficiently distant from the – and not too similar to – original works that might have been used in machine training.

In music, in particular, this distinction can be subjective — especially when AI models imitate rather than copy pre-existing works.

## Legal and Policy Developments

Around the world, lawsuits are testing the above-mentioned boundaries.

- GEMA v OpenAI and GEMA v Suno in Germany challenge unlicensed training on music recordings and song lyrics.
- In the U.S., the RIAA and major labels have filed similar claims against Suno and Udio, seeking statutory damages.
- Visual artists and media companies — from Getty Images to the New York Times — allege mass scraping of their protected materials.

These cases highlight two core issues:

1. Was the training lawful?
2. Were any opt-outs or reservations effective?

Their outcomes will shape where and how AI models are trained.

If U.S. courts expanded the U.S. doctrine of “fair use” into this domain, developers might train their machines outside of the EU and deploy them in Europe. It's worth recalling here that ‘fair use’ is a flexible legal mechanism adopted in the US - and non-existent and, therefore, non-applicable in the European Union. The US ‘fair use’ doctrine allows judges to exempt socially valuable uses of copyright works from copyright's scope. One of fair use's main requirements is that a given use should not have a disruptive or too negative impact on the main market for the protected works.

A prominent example of the application of this doctrine, which might be followed in the domain of AI, was the ruling of the U.S. Court of Appeals for the 2nd Circuit in *Authors Guild v. Google, Inc.*, No. 13-4829 (2d Cir. 2015). Here, the Court concluded that Google's large-scale digitisation of copyright books belonging to thousands of members of the Authors' Guild, and the creation of a search functionality for these works, together with the display of snippets from such works on Google's servers, was non-infringing and, therefore, fair use. What mattered for the Court was that the purpose of copying was highly transformative, and the public display of copyrighted texts was limited in a way that these excerpts did not provide a significant market substitute for the protected works.

## The EU's Answer: The AI Act — Copyright & Transparency

The AI Act adopted in 2024 adds a compliance layer for general-purpose AI, which also includes generative AI. This Act says, under its Article 53(1)(c), that AI providers shall adopt a copyright-compliance policy, particularly with regard to respecting the right-holders' opt-outs contemplated in the DSM Directive (cf. Article 4(3)).

This is an obligation that AI developers shall comply with, irrespective of where the AI training took place, meaning that even US tech companies such as OpenAI must adhere to it when training their AI systems in the United States.

The AI Act also provides (under Article 53(1)(d)) that AI developers must publish a sufficiently detailed summary of the materials used to train their models. If enforced, these obligations can help rightholders verify lawful access to their works and whether their reservations were duly considered — shifting the burden toward transparency.

The recently established EU AI Office is given the institutional task to oversee these rules and help stakeholders properly understand the AI Act's legislative requirements and their duties through the recently published Code of Practice for general-purpose AI, which has already been signed by major AI companies such as OpenAI, Google and Meta.

## Fairness, Licensing, and the Data Gap

To conclude: two policy levers are likely to determine whether and how AI and well-established music copyright can coexist:

- As regards the 'input' phase, rights reservations can signal a copyright holders' *offer to license* — using metadata or robots.txt to authorise training for a fee, possibly through collective licensing systems.
- As regards AI-created output, the AI Act's marking obligations for synthetic content (cf. Art. 50) can support and facilitate attribution and downstream remuneration when AI-generated works enter the market. However, both require reliable usage data and the auditing of this data.

Without interoperable identifiers, verifiable usage logs, and transparent metadata, even the best licensing models would fail.

In the domain of recorded music - in particular, as we have seen in the previous episode - fairness depends on data interconnections. With several layers of rights and their respective co-owners, every creative contribution can be recognised and properly remunerated only through interoperable identifiers, matched repertoires, and verifiable logs. Without this data-sharing infrastructure, fair remuneration remains out of reach.

## Conclusion

As our exploration of Digital Rights Awareness comes to an end, one truth stands out: every leap in technology forces us to redefine fairness. From the birth of authors' rights to the age of algorithms, the story of music has always been a story of adaptation: of laws catching up with innovation, and of creators fighting to stay visible in an ever-changing landscape.

Artificial intelligence may be the newest disruptor, but the question remains the same:

How do we ensure that creativity, at least human creativity, continues to be recognised, rewarded, and respected?

This is not the end of the discussion; it is only the beginning of a new conversation. We invite you to watch or revisit all the episodes in this series, and to keep exploring how technology, law, and creativity can coexist in a fair and human-centred digital future.

Because fairness, like music itself, must be constantly re-created and re-considered.