IAEL 50 YEARS

INTERNATIONAL ASSOCIATION OF ENTERTAINMENT LAWYERS

The International Music Business Navigator
for Lawyers and Artists





How European AI Legislation Affects the Music Industry Practices

By Benjamin van Werven

WALDEN GRENE

How European AI Legislation Affects the Music Industry Practices



Author: Benjamin van Werven

» 1. Introduction

In 2019, responding to "rapid technological developments" and the emergence of "new business models", the European legislator enacted the Copyright in the Digital Single Market Directive [1] (also the "DSM Directive"). Among its key innovations was an exception for commercial text and data mining (the "TDM exception"), designed to strike a fair balance between the interests of rights holders and users. [2]

This article examines the commercial text and data mining exception under Article 4 of the DSM Directive, with a focus on the preparatory phase of AI training (rather than on the outputs of generative AI systems) and considers its implications for agreements in the music industry. It also addresses the contested question of whether the training of general-purpose AI ("GPAI") models falls within the scope of the TDM exception, or whether it goes beyond it. Finally, the author aims to offer practical guidance to rights holders on how to prevent AI systems from being trained on their works.

» 2. The 'text and data mining-exception' in the DSM Directive and the AI Act

2.1 'Text and Data Mining'

The Directive defines 'text and data mining' as an *"automated analytical technique aimed at analysing text and data in digital*

Artist Question

"How can I best protect my interests in my music against use or misuse by AI?"

form in order to generate information which includes but is not limited to patterns, trends and correlations". The key characteristic is that the analysis is automated and directed at extracting new insights from existing content to train GPAI models, for example those adopted for Claude and ChatGPT. Consequently, text and data are analysed using text and data mining techniques, after which only unprotected information is extracted from the works and incorporated into the model's parameters. The original files used, and the protected information from it, must subsequently be deleted.

After the major breakthrough of GPAIs like ChatGPT in 2022, the European Artificial Intelligence Act [3] (also the "AI Act") of 2024 confirmed that the TDM exception from Article 4 of the earlier DSM Directive also applies to the text and data mining for GPAI models.[4]

2.2 'Lawfully Accessible Works'

The exception applies only to reproductions and extractions of *lawfully accessible works*. To satisfy this requirement, the GPAI Code of Practice ("COP") – signed by major providers including OpenAI, Google, Anthropic, Microsoft, and Amazon – requires signatories to refrain from circumventing technical protection measures and from crawling sites manifestly intended to enable copyright infringement. [5]

2.3 Opt-out

The exception does not apply if a rights holder has expressly reserved the right to prevent text and data mining. That reservation must be made in an "appropriate manner", such as machine-readable means for online content. [6]

To comply with the obligation to respect machine-readable rights reservations under Article 4(3) of the DSM Directive, the COP requires signatories' web crawlers to recognise robots.txt- files [7] and other appropriate machine-readable protocols as an express reservation of rights, including through the use of state-of-the-art technologies. The robots.txt protocol has a notable limitation, however: it can only block either specifically named scrapers or all crawlers indiscriminately, including those used by traditional search engines such as Google and Bing.

2.4 Retention

Reproductions and extracts generated through lawful text and data mining can be kept for as long as necessary for the text and data mining and should thereafter be deleted.

2.5 Compliance

The AI Act's obligations apply to all GPAI providers placing models on the Union market, regardless of where those models were trained. [8] From 2 August 2025, providers must comply with their obligations under the Act. Models already active on the market at that date have until 2 August 2027 to achieve compliance [9], though providers are not required to retrain or unlearn a model if doing so is technically or economically infeasible, provided that infeasibility is documented. [10]

Compliance is monitored by the European AI Office, which may impose penalties of up to 3% of a provider's global annual turnover for breach of the Act's obligations.

» 3. Practice and Case Law

3.1 Do GPAIs Fall Under the Exception, or do they go Further?

There is much debate as to whether GPAI providers can rely on the TDM exception, and whether the training of their models is in fact limited to "text and data mining" within the meaning of the Directive [11], that is, the automated analysis of data to generate information such as patterns, trends and correlations.

Several providers of GPAIs, such as Google, OpenAI, Anthropic, Meta and Microsoft [12] have maintained that their models are only trained by such analysis, and that works are not stored (or memorised) in the model (or that it is only "exceptionally rare" that it occurs). [13] However, it seems that the functioning of GPAIs remains a black box, apparently even to their providers. Various experts are convinced that reproductions of the source files, including copyright-protected materials, are actually permanently stored, and consequently (permanently) reproduced, within the 'trained' models. This would explain the phenomenon of 'memorisation'. If so, providers could not rely on the TDM exception in respect of that (permanent) reproduction, since the exception only covers reproductions in the training phase (after which the original files and protected materials must be deleted).

In early 2026, researchers from Stanford and Yale Universities found strong indications that the GPAI models they studied store training data rather than merely analysing or learning from it. Large quantities [14] of copyrighted text could be extracted from the models: Anthropic's Claude, for example, reproduced four books, including Harry Potter and the Philosopher's Stone and Nineteen Eighty-Four, with an accuracy of up to

96 percent. A subsequent study by researchers from Stony Brook University, Carnegie Mellon University, and Columbia Law School reached similar conclusions, finding that the models examined could reproduce up to 90 percent of book content, with verbatim spans exceeding 460 consecutive words. [15] While these findings provide strong evidence that memorisation occurs, the question has not been resolved with absolute certainty and remains a matter of ongoing, if perhaps narrowing, debate.

The same question has come before European national courts. Before the Munich Regional Court [16], GEMA claimed that OpenAI had not merely analysed song lyrics for ChatGPT but had memorised them, going beyond what the TDM exception permits. GEMA demonstrated this by submitting prompts requesting specific lyrics, which ChatGPT reproduced in near-identical or only slightly adapted form. Given the similarity to the originals, GEMA argued that the lyrics must have been reproduced unchanged within the model, requiring the rightsholders' permission.

The court agreed: given the complexity and length of the original lyrics and the nature of ChatGPT's output, the resemblance could not be coincidental. Memorisation had occurred, constituting a reproduction within the meaning of copyright law [17] and falling outside the scope of the TDM exception. OpenAI had therefore infringed the copyright in the lyrics.

As of this article's drafting, there is no case law from the European Court of Justice on these matters. However, in the *Like Company v Google* case, preliminary questions have been referred to the European Court of Justice that are highly relevant in this context. These include the question of whether training an LLM-based generative chatbot (being Google's Bard here (now Gemini)) by observing and matching patterns in order to enable the model to recognise linguistic patterns constitutes an act of 'reproduction', and whether this falls within the scope of the TDM exception. It also remains to be seen whether the judgment will apply to all GPAI models more broadly – particularly those employing a different technology than Google's Bard/Gemini – whose already-trained model appears to use live web content as input, processing it to produce the requested summaries.

3.2 Reservation of Rights

In a separate case before a German court, the question arose whether a rights reservation had been made in an "appropriate manner" within the meaning of Article 4(3) of the DSM Directive. The case concerned a stock photo agency that had restricted scraping and downloading of its images using "automated programmes," both in its terms of service and in the website's source code. Despite this, the scraper of LAION, an organisation that created a dataset for AI training models, had collected the images.

LAION argued that, at the time of scraping (2021), a reservation expressed in natural language and embedded in source code was not machine-readable by its scraper and therefore did not meet the "appropriate manner" requirement.

The Higher District Court of Hamburg [18] decided that it had not been demonstrated that, at the time the specific photograph was downloaded, technology existed capable of reading a reservation in natural wording within the terms and conditions of a website or its source code. The reservation had therefore not been made in the 'appropriate manner'.

Furthermore, the judgment held that a notice by a licensee (such as the photo stock agency) could also suffice, and that it is not necessary for the author to have done so himself. This also follows from the provision in the Directive that the notice must be made by a 'rightsholder', as implemented in German law as well.

The decision turned heavily on what was technically feasible in 2021; by 2026, it seems clear that reading a reservation expressed in natural language or embedded in source code is no longer beyond the capability of web crawlers. Moreover, the major GPAI providers have agreed in the COP that a robots.txt file constitutes a valid reservation.

3.3 Collective Opt-outs

In practice, reservations are made not only in relation to individual (online) publications of the works in question, but also 'collectively', irrespective of the platforms on which the works are published. For example, various collecting societies, including Buma/Stemra, SENA and SACEM have made a TDM reservation for their repertoires. Similarly, Sony Music has published a *"Declaration of AI Training Opt Out"* in which it prohibits any use of its recordings and compositions for text and data mining, scraping or similar processes, unless an explicit licence has been granted.

Some major rights holders are consolidating their reservations in a centralised database to improve visibility and discoverability. The most prominent example is *RightsAndAI* [19], launched by the International Confederation of Music Publishers (ICMP), which reports that rights holders representing over 80% of the global music publishing market (by share) have participated. The reservations made cover the entire repertoires of the rights holders rather than individual works.

For works made available online, it is questionable whether a declaration posted on a different website from the one hosting the work qualifies as an opt-out *'expressly reserved in an appropriate manner'*. The rationale of the opt-out mechanism is that scrapers can detect a reservation when scraping a given site. A reservation hosted elsewhere may therefore not function as intended. [20]

A further problem arises where such opt-out, even if technically valid in form, fails to identify the specific works it covers, whether on the same site or in a separate section of, for example, a collecting societies' repertoire database. Such reservations must also be machine-readable. In *Kneschke v. LAION*, this condition was not met because the opt-out was expressed in natural language alone, which failed to satisfy the applicable technical standards as of 2021. That requirement may be applied with even greater strictness to repertoire databases on websites.

3.4 Practice and Case Law Conclusions

At this point in time, it is unclear whether the TDM exception – intended i.e. for *analysing* but not *permanently storing* copyrighted works for the functioning of GPAIs – actually covers the practice of GPAIs. Consequently, it is also unclear at this moment whether rights holders are left empty-handed if they have not entered a reservation against TDM, or whether GPAI providers also infringe in case it uses such works for its models if no reservation has been made.

Until the 'black box' of text and data mine model training is opened, rights holders can, at best, rely on the *probability* (as accepted by the court in the *GEMA v. OpenAI* case) that their works have been memorised in the GPAI model. In any case, if rights holders wish to prevent that their works are used for GPAI models, it is recommended that they at least reserve that right and ensure that their licensees do the same.

For online publications, whether a reservation is made in the appropriate manner depends on current practice. At present, a *robots.txt*-file seems to be generally considered sufficient, including vis-à-vis GPAI providers that have not committed to the COP.

To avoid uncertainty, rights holders making reservations in respect of online publications should therefore do so via a *robots.txt* file or another generally accepted machine-readable means that is not subject to the limitation of having to block either specifically named scrapers or all crawlers indiscriminately.

Although the Directive does not specify where a rights reservation must be communicated, copyright notices should, as a matter of good practice, be placed at the point of publication rather than on a separate or third-party website.

» 4 Impact on Deals

Alongside and following legal actions, the music industry is increasingly turning to licensing agreements and partnerships with AI providers.

4.1 Licensing Agreements

Universal Music Group has concluded agreements with a range of GPAI providers, including with Stability AI, allowing the latter's models to be trained "responsibly" for its AI products on UMG's catalogue. [21] AI startup Klay Vision has also entered into a licensing agreement with UMG, as it also did with Warner Music Group, Sony Music Entertainment, and their respective publishing divisions. The aim is stated to be developing a "framework for an AI-driven music experience" built exclusively on licensed music from authors and artists who have given their explicit consent.

WMG has similarly granted a license to Udio, permitting the platform's users to create remixes, covers, and new songs using the voices of artists and compositions of songwriters that chose to participate. Participating artists and songwriters receive attribution and two separate fees: one for the use of their music in training the AI model, and one when subscribers use their songs to create new works. [22] The music created would only be shared within the platform's 'protected environment', which would prevent competition with original 'man-made' music.

Suno refuses to adopt such a 'walled garden' approach, which prompted UMG to withhold a licence and continue its legal proceedings against the company, though this approach did not prevent WMG from agreeing a licence agreement with Suno. [23]

The financial terms of these deals remain opaque, and it is unclear how and in what amounts individual works and their creators will be credited and compensated. It is neither clear whether all such similar licences also apply solely to artists that chose to participate.

4.2 Artist Agreements

The latter point is particularly important where publishers or record labels have not made explicit artist consent a condition of making works available to AI services. Artists and authors may not have full visibility into, or control over, how their material is used to train GPAI models.

Traditional recording and publishing contracts with major labels and publishers are typically worded broadly enough to cover all current and future forms of exploitation, which means they potentially also include the right to train GPAI models on those works.

Under Dutch law, transfers of copyright and exclusive licences granted by natural person authors are construed narrowly: only rights expressly stated in the agreement, or necessarily implied by its nature and scope, are deemed transferred. Non-exclusive licences may be interpreted more broadly, taking into account the parties' reasonable expectations. Separately, statutory moral rights, including the right to object to

alterations of the work, apply unless it would be unreasonable for the creator to invoke them, and may be waived, as is common in Dutch practice.

Regardless of the applicable legal system, it is advisable to include an express provision prohibiting AI-assisted editing of the works, to avoid any ambiguity on the issue of granting licenses to use music for AI.

If an author or artist does *not* wish licenses to be granted for the training of AI models on their music, it is recommended that they require their licensees to ensure that any publication of their music by a sub-licensee is accompanied by a machine-readable rights reservation, such as a *robots.txt* file or any other means that does not have the limitations inherent in it and that may be considered to be state-of-the art in time.

If an author or artist is however *willing* to agree that their music will be used for AI by third parties such as (sub)licensees, it can be inferred from the case law and practice described above that it is recommendable to include provisions in the agreement regarding how and to what extent the music from the original works may be recognisable in the final product [24]; how attribution for AI-generated music will be handled; and when and how royalties will be calculated and distributed to participating artists; whether outputs are restricted to the provider's walled garden (or other 'protected environment') and how that restriction will be enforced.

5. Conclusion

Generative artificial intelligence is viewed by many in the creative industry as a threat that exploits their works without compensation, only to compete with – or displace – the very creators whose work made it possible.

Whether we like it or not, AI is an increasingly significant part of everyday life, and all stakeholders, including legislators, GPAI operators, and rights holders alike, must find a way to engage with it.

Not only does the pace of development make this difficult, but also the opacity of how GPAI models are trained, raising the question of whether current legislation is keeping pace with practice or, as is so often the case, lagging behind it.

Either way, some lessons can be drawn from current practice, both for those who wish to restrict the use of their work by AI as far as practically possible, and for those who choose, or find themselves compelled, to embrace it.

© Copyright 2026, Benjamin van Werven

[1] *Directive 2019/790/EU.*

[2] *DSM Directive, recitals 3, 5, 6, 8.*

[3] *Regulation (EU) 2024/1689 of the European parliament and the Council of 13 June 2024.*

[4] *AI Act, recital 105 and 106 and Article 53 (1) (c).*

[5] *The COP constitutes the instrument through which these providers fulfil their AI Act obligation to adopt policies complying with the DSM Directive.*

[6] *Article 4 DSM Directive.*

[7] *A robots.txt file is a text file placed in the root directory of a website in which the operator can instruct web crawlers whether they are permitted to crawl the site or specified parts of it. The protocol originated from self-regulation within the internet sector and is supported by virtually all major search engines.*

[8] *AI Act, recital 106.*

[9] *Article 111(3) AI Act.*

[10] *Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act), C(2025) 5045.*

[11] *Article 2 (2) DSM Directive.*

[12] *www.regulations.gov/comment/COLC-2023-0006-9003, www.regulations.gov/comment/COLC-2023-0006-8906, www.regulations.gov/comment/COLC-2023-0006-9021, www.regulations.gov/comment/COLC-2023-0006-9027 and www.regulations.gov/comment/COLC-2023-0006-8750.*

[13] *As claimed by OpenAI, www.regulations.gov/comment/COLC-2023-0006-8906.*

[14] *Ahmed, Feder Cooper, Koyejo and Liang, Extracting Books from Production Language Models (2026), arxiv.org/pdf/2601.02671. The models studied were OpenAI's GPT-4.1, Google's Gemini 2.5 Pro, and Anthropic's Claude 3.7 Sonnet.*

[15] *arxiv.org/abs/2603.20957. The models studied were GPT-4o, Gemini-2.5-Pro, and DeepSeek-V3.1.*

[16] *Munich I Regional Court, 11 November 2025, case 42 O 14139/24 (GEMA v. OpenAI).*

[17] *"By any means and in any form", consistent with German case law and the CJEU's approach in Case C-476/17 (Pelham of 29 July 2019).*

[18] *Hamburg District Court 27 September 2024, 310 O.22723 (LAION v Robert Kneschke).*

[19] *rightsandai.com.*

[20] *This view is also supported in EUIPO, The Development of Generative Artificial Intelligence from a Copyright Perspective (May 2025), p. 165 and 167.*

[21] *https://www.universalmusic.com/universal-music-group-and-stability-ai-announce-strategic-alliance-to-co-develop-professional-ai-music-creation-tools/.*

[22] *https://www.digitalmusicnews.com/2025/10/29/umg-udio-deal-everything-we-know/.*

[23] *K. Robinson, 'Walled Gardens' Are at the Center of the AI Music Debate — Can (And Should) the Flood of AI Content Ever Be Stopped?, Billboard Pro 17 February 2026.*

[24] *Book publisher HarperCollins, for example, offers authors an opt-in scheme under which outputs are capped at 200 consecutive words or 5% of a book's text (https://authorsguild.org/news/harpercollins-ai-licensing-deal/?idp=https%3A%2F%2Fengine.surfconext.nl%2Fauthentication%2Fidp%2Fmetadata).*