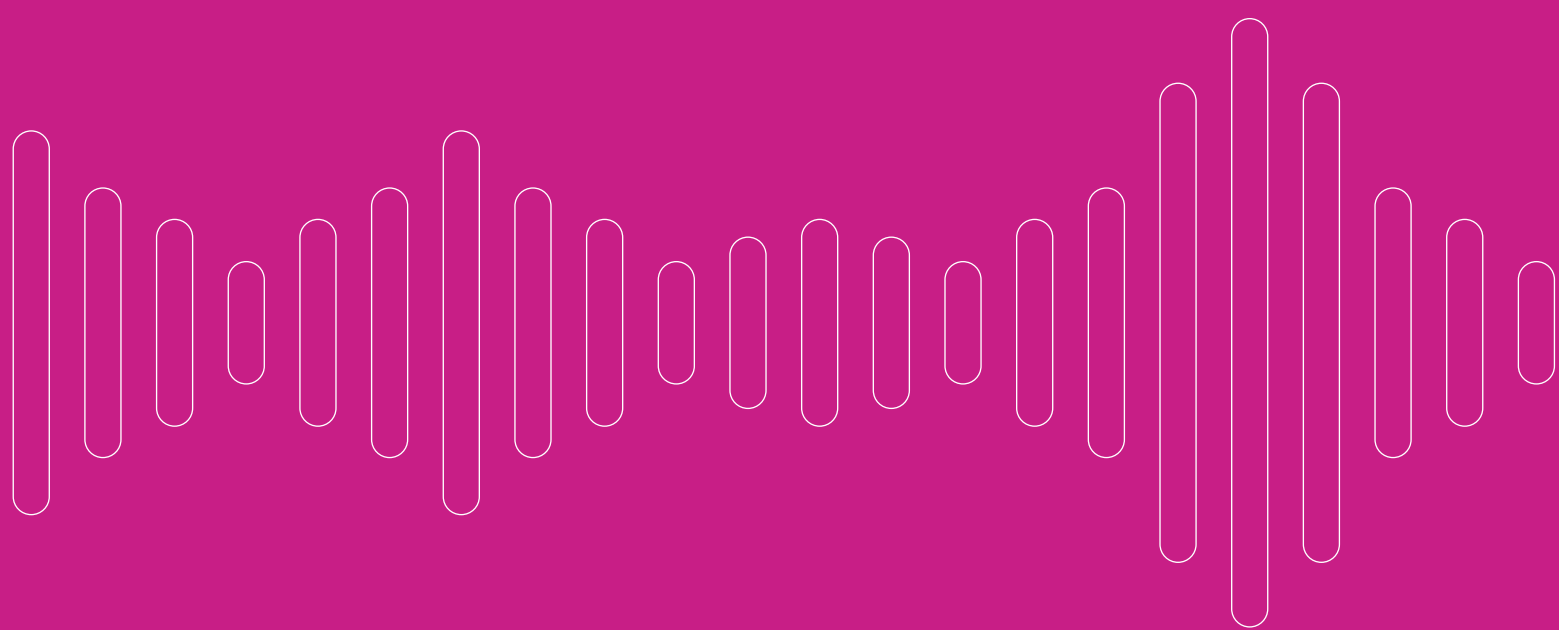


GenAle

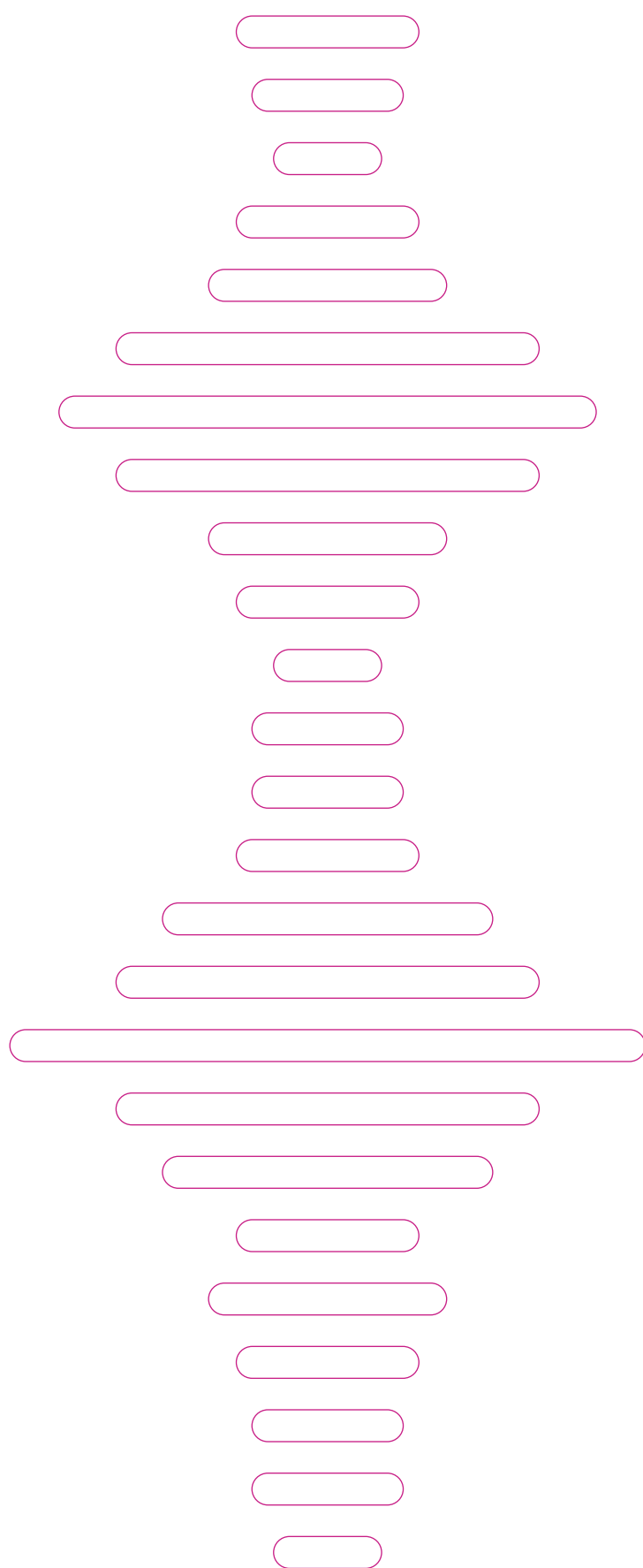
Text and Data Mining as a Possible Exception to Intellectual Property Rights in Generative AI

April 2026



GenAle

Powered by 



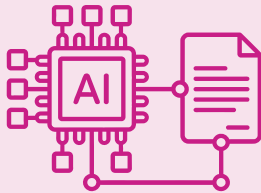
This work and all its contents are expressly protected by intellectual property rights, owned by AIE, Artistas Intérpretes o Ejecutantes, Sociedad de Gestión de España. Reproduction and public dissemination of this work is permitted as long as there is no commercial or economic purposes, provided that the source is properly cited. Its physical distribution and transformation is explicitly prohibited without the express prior authorization of AIE. If you wish to exercise any of the rights not recognised in this clause, you can contact AIE through the following email:contacto@genaie.es.

To cite this publication, use the following bibliographic reference:
AIE and GenAIE (2026): Text and Data Mining as a Possible Exception to Intellectual Property Rights in Generative AI.

Table of Contents

Executive summary	4
 Text and data mining, artificial intelligence and intellectual property	6
Legal definition of “text and data mining”	6
Relationship between text and data mining and Artificial Intelligence	7
Justification of TDM in current regulations and its relevance for law	8
 The dual exception for TDM. Requirements	12
Background and justification of TDM’s dual exception	12
The TDM exception for scientific research	13
<i>Nature of the exception and Spanish transposition</i>	<i>14</i>
<i>Beneficiaries</i>	<i>15</i>
<i>The exception’s content and material scope</i>	<i>16</i>
<i>Inexistencia de compensación por la excepción</i>	<i>17</i>
The TDM general exception allowing commercial uses	17
<i>Nature of the exception and Spanish transposition</i>	<i>19</i>
<i>Beneficiaries</i>	<i>19</i>
<i>The exception’s content and material scope. The role of opt-out</i>	<i>19</i>
Comparison Table: art. 3 vs art. 4 of the DSMD	26
 Differences with the U.S. approach (fair use and copyright market dilution)	27
 Recommendations on the TDM exception and its evolution in generative AI	31

Executive summary



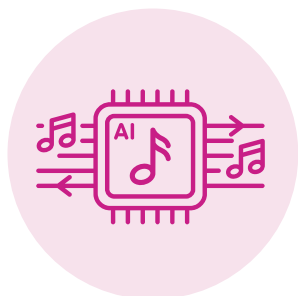
“Text and Data Mining” (TDM) is, in technology’s current state, essential for the functioning and training of Generative Artificial Intelligence (Gen AI), since it allows automatic analysis of large volumes of digital information in order to identify patterns and correlations, which requires making certain copies of works that are potentially protected under copyright. Therefore, its practice has direct implications for regulation of intellectual property and requires a specific legal framework.



Directive (EU) 2019/790 establishes two mandatory exceptions for TDM.

The first, for scientific research (art. 3), allows public agencies and non-profit cultural organisations to do TDM with lawful access and to keep the necessary copies for verification of results.

The second, of a general nature (art. 4), allows TDM for any purpose, even commercial, but subject to the rightsholders not having activated an explicit, appropriate and machine-readable opt-out. This exception only covers reproductions strictly necessary for analysis and prohibits retaining copies beyond the essential technical time.



That a model generates substantial fragments from the material it was trained with (works and/or subject matter) demonstrates that it has retained them, going beyond what the exception allows, which covers only analytical operations, not “memorisation”.

The current regulation, designed to exclude the use of exclusive rights merely for analytical purposes, is insufficient to balance the interests between AI developers and rightsholders, and European reforms are already being proposed that require greater transparency, effective opt-out systems (or, directly, opt-in systems) and new licensing models for Gen AI.

In this context, as a result of the analysis done, GenAle, AIE's Analysis Unit on AI and Music in the Digital Era, considers it necessary to advance a structural review of the TDM system applicable to generative AI, aimed at ensuring an effective balance between technological innovation and protection of intellectual property rights. Specifically, it is suggested to:

1



Adapt the current regulatory model, overcoming its original limitations and moving toward prior authorisation frameworks (opt-in) for certain cases.

2



Strengthen obligations for transparency, control and effective respect for the reservation of rights, as well as clarification of the legality of the datasets used.

3



Promote the development of new licensing frameworks, , including collective solutions (such as extended collective licensing) that make it possible to reconcile innovation and fair remuneration for rightsholders, along with consideration of compensation mechanisms for past uses.

4



Progress in the legal protection of performers, specifically in regard to the use of their voice and image in generative AI environments.

TEXT AND DATA MINING, ARTIFICIAL INTELLIGENCE AND INTELLECTUAL PROPERTY

Legal definition of “text and data mining”

Art. 2.2 of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (referred to as the DSMD throughout this text) defines “text and data mining” (TDM) as:

“Any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations”.

Spanish legislation includes this definition in art. 66.1 of Royal Decree-Law 24/2021 of 2 November¹, with some inconsequential textual variation:

“Any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes patterns, trends, correlation and similar elements”.

The relevant thing about text and data mining is the “automated analysis”² of information (all types of data that are provided to the system, subject to intellectual property or not) and the purpose of same, which is the generation of patterns, trends and correlations. Even when this purpose does not seem closed (“among others”), it does not cover (cannot, because it exceeds the concept itself of analysis) the reproduction or copying as output of this type of input information. This is particularly relevant when such information or data is considered to be a work or rendering protected by Intellectual Property Law (IP).

¹ Royal Decree-Law 24/2021 of 2 November, transposing European Union directives in the areas of covered bonds, cross-border distribution of collective investment entities, open data and reuse of public sector information, exercise of copyrights and related rights applicable to certain online transmissions and retransmissions of radio and television programs, temporary exemptions for certain imports and supplies, for consumers and for the promotion of clean and energy efficient road transport vehicles.

² The designation of the exception as “mining” can be considered reductionist. Following French doctrine, JIMÉNEZ SERRANÍA highlights that this term refers to data extraction from various information resources. In contrast, the concept of “analysis” (which is used to define mining: “analytical technique ... intended to analyse...”) is broader, encompassing many more actions: to track, process, compare, retrieve, interpret, search, classify, delete, etc. (JIMÉNEZ SERRANÍA, 2020, p. 248).

Relationship between text and data mining and Artificial Intelligence

An Artificial Intelligence system (AI), as defined by the European *Artificial Intelligence Act*³ (AIA, art. 3(1)), can infer, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions. *Generative AI* (Gen AI) is a general AI system that allows flexible generation of content in text, audio, images or video format⁴. Thus, Gen AI can provide novel results from the patterns and correlations “learned” from the inputs provided and based on the instructions given to it, or regurgitate it as is, also as a result of such instructions or prompts, fragments of information or already existing complete information, including protected works and renderings that have been stored during the training process.

It is essential, then, to determine what AI models really do: they do not “read” the human way (nor “enjoy” any content), but rather extract patterns and correlations from the data or information presented to them. In doing so, or rather, *in order to do so*, they need, in the current state of the technology, to reproduce, during their learning process, the information (upload it into the system’s memory), even if only temporarily, while analysing it.

Text and data mining is necessary to train Gen AI systems, such as Large Language Models (LLMs), which require huge amounts of information. To obtain these *corpora* or *datasets* it is necessary to extract, compile and process texts or data on a large scale.

TDM techniques convert raw data (unprocessed, duplicated, noisy, unreadable...) into useful data for machine learning. Without TDM (mass data extraction and processing) there is no *corpora*, without *corpora* or *datasets* there is no training, and without training there is no AI (or Gen AI) model. During the training process, own or third-party datasets may be used, but **the creation of these datasets already constitutes text and data mining**, as it involves, from a technical point of view, performing data scraping activities (crawling, scraping), syntactic analysis (parsing), deduplication, filtering, standardisation and *tokenization*.

3 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).

4 Recital 99 of the AIA points to Gen AI as a typical example for a general-purpose AI model, given that “it allows for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks”. Art. 3(63) of the AIA defines the “general-purpose AI model” as “an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market”.

Justification of TDM in current regulations and its relevance for law

This technical processing of data involves, from the perspective of copyright law, the performance of legally relevant acts, from simple extraction (of importance, for example, in regard to databases) to the making of copies or authentic “reproductions” in the legal sense.

The law takes into account these processing operations by accommodating a concept within the regulation that now goes beyond the technological to have legal content: text and data mining.

Within the framework of its **Digital Single Market Strategy** (COM(2015) 192 final), the European Commission carried out a technical analysis, included in its **Staff Working Document SWD(2015) 100 final (Analysis and Evidence)**, which, among other issues, sought to identify the structural barriers preventing development of the European digital economy. Although it did not end up proposing regulatory solutions, this document collected evidence that, one year later, would lead to formulating the first proposals to reform copyright law, recognizing the exceptions for TDM currently in force.

In *Document SWD(2015) 100 final*, the Commission recognized that modern scientific research is increasingly dependent on access to large volumes of data, which, in Europe, is hampered by *legal obstacles* (lack of legal clarity: researchers did not know whether TDM infringed intellectual property rights, as it might involve reproduction of content) and *restrictive contractual practices* (rightsholders tended to prohibit or limit automated uses through licenses), *along with territorial fragmentation of markets* (each Member State could apply different rules, and contractual licenses were also granted according to State scope). These limitations affected not only cultural consumption, but also innovative uses that required processing large volumes of content, such as text and data mining (although this expression is not always used), and cross-border projects that relied on automated data analysis. As a result, Europe ran the risk of losing competitiveness when faced with international competitors having more flexible and permissive legislation on mass data analysis (the US, above all, operating under the fair use paradigm).

This analysis constitutes the empirical and conceptual basis that will lead the Commission, one year later, to propose the introduction of mandatory exceptions for TDM in the proposal for a Copyright Directive in the Digital Single Market (COM(2016) 593 final), that would finally materialize in current articles 3 and 4 of the DSMD in 2019. The Impact Assessment accompanying this proposal for a Directive (SWD(2016) 301 final and SWD(2016) 302 final) emphasized the competitive disadvantage and the need to regulate TDM to allow European researchers, in particular, but also companies, a freer and safer recourse to these techniques.

As can be seen, the regulation was designed with academic and scientific *Big Data* in mind (particularly appreciable in the exception of art. 3), **before the roll out of technology and business models around commercial Gen AI** (LLMs, music and video generation models, etc.). Even with its extension to the commercial use of TDM (art. 4), what the Directive sought to legalize was “simple” analysis, *analytical reading*, not *generative reproduction*.

It is not the only legislative approach to mass data analysis techniques. *Scraping technology* (scraping or extracting data from one, several, or hundreds of websites), which enables new business models in the modern Data Economy, is also covered by the database regulations (and the case law of the CJEU that develops it, around metasearch engines and comparators) and art. 15 of the DSMD, in relation to news aggregators, even without qualifying said technology in either case.

Indeed, scraping (scraping or extracting data from the web) is nothing more than a technical concept, which encompasses operations that may or may not be part of a TDM process, in the same way that not all mining requires scraping⁵. “Text and data mining”, however, transcends the mere technological consideration to become a *legal category*, having been taken into consideration by the legislator, who defines and regulates it, even as an exception to the exclusive rights of holders of protected works and subject matter⁶.

Although AI in practice needs both, only TDM is regulated as such.

In this context, the (dual) exception of TDM *in terms of Intellectual Property* was designed to remove obstacles to the use of mass data analysis techniques necessary for the development of science and European markets, so that neither the States (through territorial regulations) nor exclusive rightsholders (through their lawful exercise) could indirectly prevent achieving a globally competitive European Digital Single Market, where prior knowledge could be used for new research and business models⁷.

In the European-Continental tradition of copyright law, rightsholders are granted a series of exclusive rights (moral and economic) that allow them normal exploitation of protected works and subject matter. These rights, however, are not absolute: In order to guarantee other interests worthy of protection, the legislator sets limits or exceptions to them, with or without compensation for the aforementioned rightsholders. Using such exceptions, third parties may perform acts they would not normally be authorised to do.

Text and data mining is regulated based on this consideration, and therefore involves a necessary

5 Being a previous phase of mining, *not all scraping is TDM* (the collection of product prices, extraction of emails or contacts, downloading images without their analysis, or copying content for republication are not), *nor does all TDM require scraping*: thus, when an already licensed *corpus* is analysed, internal databases are processed or public datasets are worked on that have been legally accessed.

6 There are also some fundamental differences between a purely *syntactic scraping* and the *semantic/analytical scraping* typical of TDM: the first is essentially a “reading for copy/paste” model (data is extracted for reproduction in a new context), while the TDM model and its exception is “reading for analysis/understanding/learning”. Thus, for example, a comparator can do syntactic scraping of thousands of websites to provide specific information (the price of a flight, or of many) that is shown as is to the user; in doing so, it competes with (replaces) the original source, which the user will not go to. In the case of semantic scraping, typical of TDM, whoever does it scans these thousands of websites not for ticket sales, but rather to predict price trends (pure analytics). In this case there is no market substitution whatsoever.

The TDM exception is not designed to protect syntactic scraping, because it does not intend to allow replacement of the original work or to create a directly competing product that reuses the creative content, a conclusion that can be reached by relating recital 18, arts. 3 and 4 and 7.2 of the DSMD and art. 5.5 in the Directive on Copyright in the Information Society: As TDM is conceptually limited to the mere extraction of patterns, the exception provided for in the 2019 Directive makes the validity of the exception (art. 7.2) conditional to strict compliance with the rule of the three steps enshrined in art. 5.5 of the Directive on Copyright in the Information Society, thus preventing TDM from serving as legal cover for the development of generative models or services that conflict with normal exploitation of the work or subject matter and/or unjustifiably harm the legitimate interests of the right holder.

This is why activities such as comparators or aggregators must base their legality on precepts that are different from arts. 3 and 4 of the DSMD.

7 See later the alternatives considered by the legislator.

modification of Intellectual Property regulations: being an exception to the holders' exclusive rights, provided that the legally intended conditions are met.

If in the beginning (and, above all, by the industry and AI developers) it could be thought that the training of an AI system or model could perhaps fit into the definition of TDM (simple extraction of syntactic and semantic rules) and, as such, have its functioning in all its phases (from the collection of inputs to the outputs) be legally protected by agreement on this exception, the truth is that the ability of Gen AI models to reproduce literal or substantial extracts of the works with which they were trained was quickly revealed, which implied an effect on exclusive rights in terms of copyright beyond what was foreseen by the law. That is, beyond, most likely, what is permitted by the exception itself.

As pointed out by the recent German judicial doctrine (*GEMA vs. OpenAI*, 2025; see later), if a model can generate the original work *as an output*, we are not facing a mere extraction of metadata or trends (object of TDM), but rather an exploitation that exceeds the scope of analytics to enter, fully, into the realm of reproduction, first, and public communication, later. But not only from the output side: it is also inferred from the case that if the input as is can be generated as an output, it means that the model has not only analysed the data supplied to extract patterns with which to respond to later requests, but has "memorised" it (stored, copied...) during this process.

That is, even the simple training of an AI system with protected works or subject matter, its mere *reading* by the model, could already imply an authentic "reproduction" for legal purposes (it meets the requirements of art. 18 of the revised text of the Intellectual Property Act, the TRLPI in its Spanish acronym) irrespective of whether or not such works or subject matter are later reproduced in the outputs.

Apart from TDM and its very limited framework of action, there is no technological exception, for legal purposes, that protects the performance of these tasks that involve reproduction of protected works and subject matter⁸. As will be seen, the exception for TDM does not even allow free analysis of data by anybody or in whatever context, but rather is subject to very specific limits that imply, on the part of the legislator, to recognize the aforementioned effect on exclusive rights from the perspective of providing input data for system training.

In short, TDM is not an end in itself, but a means⁹. The relevance for copyright law is that, in order for the software to be able to analyse these "patterns or correlations", it inevitably requires acts of reproduction (copies) of the protected works and subject matter, which, in the absence of authorisation or legal exception (that for TDM, contained in arts. 3 and 4 of the DSMD), would constitute an infringement of art. 2 of the Directive on Copyright in the Information Society¹⁰. This

8 In the historical context, the very loading in RAM memory of computer programs, ephemeral and technologically necessary, is considered reproduction for legal purposes (see art. 99 TRLPI). From the technological point of view, only the exception of network transmission is contemplated by the legislator, even recognizing its consideration as "reproduction" (art. 31.1 TRLPI).

9 Consideration 8 of the DSMD: "*New technologies enable automated computational analysis of information in digital format, such as text, sounds, images or data, which is generally referred to as 'text and data mining'. Text and data mining enables the processing of large amounts of information in order to acquire new knowledge and discover new trends*".

10 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

Recital 8 of the DSMD: "*In certain instances, text and data mining can involve acts protected by copyright, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are normalised in the process of text and data mining. Where no exception or limitation applies, an authorisation to undertake such acts is required from rightsholders*".

may be evident, but not only, when the results provide fragments of protected works or subject matter, or others, or complete subject matter, which may also result in additional infringements of exclusive rights.

THE DUAL EXCEPTION FOR TDM. REQUIREMENTS

Background and justification of TDM's dual exception

To regulate TDM in terms of Intellectual Property, the EU legislator considered three different alternatives¹¹:

- “Option 0”: maintain the status quo, the existing framework, relying on contractual solutions and the law hitherto in force.
- “Option 1” (limited exception): create a restrictive mandatory exception for TDM, only for research organisations and cultural heritage institutions, for scientific purposes.
- “Option 2” (broad exception): allow an open exception, including for commercial uses, from which any user with lawful access to the work or subject matter could benefit, except for expressly reserved rights by the owners (opt-out).

The European Commission ruled out the status quo because it kept intact the structural barriers that prevented the use of TDM in Europe: legal uncertainty, regulatory fragmentation, and reliance on restrictive contractual licenses¹².

The option of an exception limited only to research organisations was also unsatisfactory, as it would have left out private actors essential for data-driven innovation, thus perpetuating the market failures identified.

Conversely, a general exception without limitations created strong opposition from rightsholders, who feared losing control over commercial uses of their content.

The solution finally adopted — the combination of options 1 and 2 in current arts. 3 and 4 of the DSMD — represented a balance between these interests¹³: it ensured an inalienable right for public research, while allowing TDM for any user with lawful access, preserving the ability of rightsholders to keep commercial uses for themselves through an opt-out mechanism. This hybrid or dual-exception formula¹⁴ was considered by the Commission to be the only one capable of correcting the failures in the status quo, encouraging innovation, and maintaining the protection of rightsholders within a harmonised framework at the European level.

11 Impact Assessment of the DSMD *Proposal*, SWD(2016) 301 final, Part 1/3, Section 6.2.2 (“Policy options”), pp. 147–150.

12 Already noted in the above-mentioned Staff Working Document SWD(2015) 100 final, and repeated in the Impact Assessment, Section 6.2.1 (“Problem definition”).

13 DSMD proposal (COM(2016) 593 final), Explanatory Memorandum, explaining the proposal for two additional exceptions (art. 3 and art. 4). See also recitals 8–18 of the DSMD, in particular 12, 17 and 18.

14 Depending on context, “double layer” or “alternative exceptions” can also be used.

The TDM exception for scientific research

It is reflected in art. 3 of the DSMD, and specifically referred to therein, following the general introduction to TDM in recitals 8-12, in recitals 12-17 of same:

Article 3. Text and data mining for scientific research purposes

1. Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.
2. Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results.
3. Rightsholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.
4. Member States shall encourage rightsholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively.

In essence, this art. 3 of the DSMD provides for a **mandatory exception**: it contains a compulsory and unifying mandate, whereby Member States **must** establish an exception allowing **research organisations** and **cultural heritage institutions** to make reproductions and extractions of works to which they have **lawful access**, for the sole purpose of doing **TDM for scientific research purposes**.

In Spanish law, the rule was transposed, mixing in successive sections of DSMD provisions from art. 3 and art. 4, and adding the equivalent for databases and computer programs, in art. 67 of Royal Decree-Law 24/2021 of 2 November.

Nature of the exception and Spanish transposition

Article 3 of the DSMD establishes a **mandatory** exception (Member States are obliged to transpose it, they have no choice), which **cannot be blocked by contractual means** by IP rightsholders (any clause against, imposed or negotiated would be null and void, and it is not possible for rightsholders to opt-out, regulated by art. 4)¹⁵, has **finality** (only for scientific research, being excluded any commercial use, which, where appropriate, should be made in accordance with art. 4) and is **limited** (only for the beneficiaries covered).

Characterised in this way, the legislator intends, with this exception, to guarantee a sort of **absolute “safe harbour” for open European science**, allowing it to compete on an equal footing with other global actors. It is also a way of promoting open, transnational science, with reproducible results (moving from the Data Economy to the Knowledge Economy), and even laying the foundations for responsible use of AI in the public sector.

It should be noted that, with regard to national transposition, the Spanish legislator makes certain mistakes by consolidating arts. 3 and 4, making corrective interpretation necessary in accordance with the transposed Directive.

Art. 67.1 RD-L 24/2021 sets out the general principle of permitting TDM without the need for authorisation from rightsholders, understanding that it is common to DSMD arts. 3 and 4: *“Authorization from the intellectual property rightsholder will not be required for reproductions of works and other legally accessible subject matter made for the purposes of text and data mining”*.

It also considers retention obligations common (art. 67.2, although later it is qualified in art. 67.4).

Next, art. 67.3 RD-L 24/2021 states that *“the provisions of section 1 shall not apply where the rightsholders have expressly reserved use of the works to machine-readable or other appropriate means”*. This seems to allow IP rightsholders to also opt-out in regard to research organisations and cultural heritage institutions that do TDM, which is *not acceptable*, contrary to the Directive (which only permits this in art. 4, in relation to the commercial use exception).

The Spanish regulation should be read in the same sense as the European one, and considers that section 2 of art. 67 RD-L 24/2021 does not apply to TDM done by research organisations and cultural heritage institutions for the purpose of scientific research.

¹⁵ This neutralizes restrictive clauses in publishing licenses that historically blocked TDM. They are referred to in DSMD recital 10: “where researchers have lawful access to content, for example through subscriptions to publications or open access licences, the terms of the licences could exclude text and data mining”.

Specifically, in the case of *Kneschke vs. LAION* (see later) it is made clear that the opt-out is only possible against commercial TDM practices (art. 4 of the DSMD and § 44 UrhG), not scientific ones by non-profit organisations (art. 3 of the DSMD and § 60 UrhG).

Beneficiarios

Article 3 of the DSMD limits the exception exclusively to two categories of beneficiaries, both of an institutional nature related to research and culture.

- **Research organisations:** universities, public research centres, non-profit organisations or those with public interest missions.
- **Cultural heritage institutions:** libraries, archives, museums, film libraries, etc.

The Directive requires these entities to act **not for profit** (or reinvest the benefits in their scientific research) or in the framework of a **public interest mission**¹⁶.

Technically, private organisations are not excluded, provided they act with such a mission or without a commercial interest. In this regard, the German courts in *Kneschke v. LAION*¹⁷ have already used the definitions and requirements of the DSMD: Even if there is a public-private partnership, or the entity is financed by private companies (via donation of funds), or results of the Gen AI could be used by entities that are indeed for profit, none of these circumstances in themselves would be an obstacle to the exception being applied, provided that *the person who finances, or takes advantage of the result, does not exert a level of decisive influence over the non-commercial entity developing the AI, nor does this person have preferential access to the results of this research* (in this case, the dataset from LAION has general and free access).

¹⁶ Art. 2(1) and (3) of the DSMD, defining “research organisation” and “cultural heritage institution”. See also recitals 11 and 12.

¹⁷ In April 2023, photographer Robert Kneschke sued LAION e.V. before the Hamburg Regional Court for alleged illicit use of one of his creations from the stock agency (Bigstock) through which he sold them. He argued that the defendant had used an image of his without authorisation, even against his express prohibition (exercising the opt-out from art. 4 of the DSMD and the German transposition regulation, § 44 UrhG), and requested removal of the image from the LAION dataset, compensation for unauthorised use, and for the conduct to not be repeated.

LAION stands for *Large-scale Artificial Intelligence Open Network*. LAION e.V. (LAION *eingetragener Verein*, registered association) is a non-profit organisation created to develop and publish in open code, machine learning models and datasets, so they are available and reusable by third parties without the need to train their models from scratch, promoting efficient energy use and consuming less resources. Images from the LAION-5 data set, which contained the plaintiff's image, were obtained, among others, from a US set (Common Crawl), that also contained its description in text, in order to verify that they corresponded with its visible content.

In first instance, the Court of Hamburg gave its judgment (*Landgericht Hamburg, Judgment of 27 September 2024*, No. 310 O 227/23), ruling for the defendant, which is why this case is often presented as a victory for the developers of Gen AI, in the sense of allowing them to use TDM. However, the claim is nuanced, because it would be only, and in application of DSMD art. 3 (§ 60d UrhG), in relation to these non-profit uses and purposes. In the case, the Court found that LAION operated without commercial purposes and in the public interest (paragraph 110 et seq.).

The judgment was appealed, and was [upheld by the High Court](#) (*Oberlandesgericht of Hamburg, Judgment of 10 December 2025* (Az. 5 U 104/24)), although with some important nuances (actually, corrections) in regard to using the opt-out (see later). Of interest here is that the High Court ratifies that what needs to be analysed for the TDM exception contained in § 60d UrhG (art. 3 of the DSMD) to operate, versus what cannot be opposed (opt-out is not possible), says the Court, is whether or not the organisation doing the research that does the mining pursues commercial purposes or not, not whether the research also benefits third parties for commercial purposes (p. 31).

The exception's content and material scope

The exception authorises beneficiaries to do:

- **Extraction of texts or data.** The concept of “extraction” is not defined. By analogy, the one used in the database regulations can be used, so that, brought to these more general terms, it would refer to the permanent or temporary transfer of content from its original digital source (such as the database; it could be a website) to another medium, whatever the medium used or the way in which it is done.
- **Reproduction of texts and data:** In this case, the concept is the one used on a general level in the IP regulations, when such texts and/or data constitute the form of expression of a protected work or subject matter: the direct or indirect, provisional or permanent fixation, by any medium and in any form, of all or part of the work, which allows its communication or to obtain copies (art. 18 TRLPI)¹⁸.

However, the authorisation granted to beneficiaries to extract or reproduce protected works and/or subject matter is not absolute, and is subject to specific requirements, of a cumulative nature:

1. The beneficiary has **lawful access to the works**: this includes having done so through subscription, institutional license, open access, etc. If the works are behind payment or registration walls and these are not done, the access and subsequent use of TDM on them would be unlawful under this exception.

In addition, lawful access implies that the source from which this data is collected is also lawful: Sometimes, the Gen AI is trained with datasets that, despite being freely available, are themselves the result of illegal reproduction of protected works and/or subject matter.

2. Use of TDM is limited to **scientific research** purposes. The term should be interpreted broadly, clarifying recital 12 which includes both the natural sciences and the human sciences. Notwithstanding the fact that science has commercial applications, it would be opposed to activities done directly with a commercial goal, specifically aimed at launching (exploiting) products on the market. It should be taken into account that beneficiaries of the exception cannot be for profit.
3. Reproductions (copies) of the works done must be **kept only for the purposes of said research** (and, in particular, their verification)¹⁹ **and must be stored safely**²⁰. In this sense, rightsholders have the ability to control this secure storage but without preventing it or being able to avoid

¹⁸ Royal Legislative Decree 1/1996 of 12 April, which approves the Revised Text of the Intellectual Property Act, regularising, clarifying and harmonising current legal provisions on the matter.

These reproductions, which are considered as such for legal purposes (art. 18 TRLPI) are not exceptions for the technical reasons in art. 31.1 of the TRLPI, as they do not refer to technological processes necessary for data transmissions in networks. Their exception in order to allow TDM, therefore, must go through this other explicitly contemplated route.

¹⁹ The possibility of verifying results is part of the scientific method. Recital 15 talks about “scientific peer review and joint research”.

From another point of view, verifiability contributes one of the foundations of open science.

²⁰ Paragraphs 2 and 4 of RD-I 24/2021 add that reproductions, extractions and conservation must be done with full respect for the principles of legality and regulations on the protection of personal data and guarantee of digital rights. In addition, in application of provisions in the final point of art. 3 of the DSMD, it is noted that the codes of conduct signed among all stakeholders are the appropriate tools to reflect the best practices applicable so the Administration can promote their development. Along the same lines, recital 15 of the DSMD: Member States should be free to decide on relevant measures at the national level and “after discussions with relevant stakeholders”, including “the ability to appoint trusted bodies for the purpose of storing such copies”.

the use of mining.

Given that mining itself, as a legal concept, seeks only the extraction of patterns, trends and/or correlations, it can be concluded that art. 3 of the DSMD covers the automated processing of protected works and subject matter that have been lawfully accessed with the intention of analysing them, as well as the technical act of their copying (legally qualifying as “reproduction”) that is necessary for such analysis and, at best, to later compare and contrast the results of the research.

It does not, however, permit the fixation, distribution or public communication (including the method of making available to the public by wired or wireless means that give access at any time and from any place) of such works as a result of the use of Gen AI, beyond what is authorised by the limit of the citation (art. 32.1 TRLPI), if applicable.

No compensation for the exception

Nothing is said in the Directive’s articles about possible compensation or payouts to rightsholders for the use made of their works and/or subject matter under this exception. The reason is explained by recital 17 of the regulation:

“In view of the nature and scope of the exception, which is limited to entities carrying out scientific research, any potential harm created to rightsholders through this exception would be minimal”. Member States should, therefore, not provide for compensation for rightsholders as regards uses under the text and data mining exceptions introduced by this Directive”.

The TDM general exception allowing commercial uses²¹

Considered in art. 4 of the DSMD and, given the general recitals on TDM in recitals 8-12, it is expressly based on recital 18²².

21 While art. 3 only speaks about “exception”, art. 4 refers to an “exception or limitation”. The use of this expression is not clear, and whether, as seems logical, it should refer to different possibilities to be chosen by Member State legislators. The CJEU, in the case VG Wort (CJEU of 27 June 2013 in Joined Cases C-457/11 to C-460/11) *tried* to differentiate between both terms, in regard to art. 5 of Directive 2001/29: “[such] exclusive right can be, depending on circumstances, or completely eliminated, as an exception, or only limited. It is not excluded that such a limitation could imply, partially, based on the different specific situations regulated, an exclusion, a restriction, or even the preservation of said right”.

In this White Paper we use the term “exception” as it is the most commonly used when speaking about TDM in these terms, even when the Spanish legislator (with no intention to qualify anything and, once again, mixing) speaks about (cursive is ours) “the new limits on text and data mining for both scientific and commercial use” (EM RD-I 24/2021, paragraph VI, and header of Title II).

22 Recital 18 of the DSMD: “In addition to their significance in the context of scientific research, text and data mining techniques are widely used both by private and public entities to analyse large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies. Rightsholders should remain able to license the uses of their works or other subject matter falling outside the scope of the mandatory exception provided for in this Directive for text and data mining for the purposes of scientific research and of the existing exceptions and limitations provided for in Directive 2001/29/EC. At the same time, consideration should be given to the fact that users of text and data mining could be faced with legal uncertainty as to whether reproductions and extractions made for the purposes of text and data mining can be carried out on lawfully accessed works or other subject matter, in particular when the reproductions or extractions made for the purposes of the technical process do not fulfil all the conditions of the existing exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC. In order to provide for more legal certainty in such cases and to encourage innovation also in the private sector, this Directive should provide, under certain conditions, for an exception or limitation for reproductions and extractions of works or other subject matter, for the purposes of text and data mining, and allow the copies made to be retained for as long as is necessary for those text and data mining purposes (...)”.

Article 4. Excepción o limitación relativa a la minería de textos y datos

1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.
2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.
3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.
4. This Article shall not affect the application of Article 3 of this Directive.

Art. 4 of the DSMD establishes a second general **mandatory exception** that allows beneficiaries to perform **reproductions and extractions** of protected works for the purpose of applying TDM techniques, provided that the user has **lawful access** to the content. Unlike art. 3, this broad exception **is not limited to scientific research** nor public institutions: it applies to **any user**, including commercial actors, unless the rightsholders have **expressly reserved** the rights in an appropriate manner.

This last condition—the **possibility of opt-out** by the rightsholders—is the defining characteristic of art. 4, and the element introduced by the legislator in order to achieve the sought-after balance of interests between the promotion of commercial innovation and creation of European startups and technology companies and the rights of IP rightsholders.

This is the **key exception for commercial Gen AI**, under which the different models of Gen AI that operate in the market for profit, whether with free or paid versions, are attempting to operate and develop.

Nature of the exception and Spanish transposition

Art. 4 of the DSMD again contains a **mandatory** exception (Member States are obliged to transpose it, this is not optional, in order not to fragment the internal market), but with significant differences from the exception in art. 3:

- This is an **open** exception, not limited based on the type of user of the TDM or the purpose of its execution.
- It is an exception that is **conditional** on the non-existence of reserved rights by the rightsholders who, therefore, can prevent its use.

In the transposition made in art. 67 RD-I 24/2021, this exception would be included in the first three sections, bearing in mind that the third, as mentioned above, is applied exclusively, despite its open wording.

Beneficiaries

The exception contained in art. 4 of the DSMD **does not limit the categories of beneficiaries**, it is universal in scope. Thus, TDM could be done by individual persons or legal entities, technological startup companies, Big Tech, public or private institutions, for profit or non-profit, etc.

In regard to the beneficiaries, the only requirement stipulated, common to both exceptions, is that of *having had lawful access to the protected work and/or subject matter*.

The exception's content and material scope. The role of opt-out

The general exception of art. 4 of the DSMD authorises the beneficiaries to **extract** and **reproduce** text or data that are also protected works and/or subject matter, (understanding both terms the same as has been stated above in relation to art. 3), provided that:

1. The user has **lawful access** (subscription, license, open access, etc.). It should be interpreted in the same terms as those for art. 3. Nevertheless, practically all commercial models are suspected of having trained, at least in part, with datasets or websites of extremely doubtful legality²³. TDM on illegal datasets (if this condition is confirmed) or content found openly but clearly derived from acts of piracy should not be considered to meet this requirement.
2. Use is limited to authentic **TDM**, that is, analysis of works and/or subject matter to obtain patterns, correlations, trends, etc. **The purpose for which it is done is irrelevant** (scientific, personal or purely commercial), but the framework of what is permitted is closely linked to the legal definition of "text and data mining", as evidenced by the narrower scope of reproduction, which excludes extensive retention.
3. The rightsholders **have not reserved their rights** appropriately.

²³ Datasets such as Bools3, Common Crawl or The Pile are suspected of resorting to illegal scrapings. And there are pages like those from the activists in the group Anna's Archive that illegally "release" protected content on the Internet: after hacking Spotify they have begun to release almost 3 million tracks, 6 terabytes of audio, and [they systematically ignore all lawsuits and court orders to stop their activity](#).

In short, the exception covers the act of processing works to extract patterns, trends or correlations, as well as reproductions that are technically necessary for this, but not others, regardless of the sector or purpose, and provided that the rightsholders have not excluded their works from the exception by means of an appropriate opt-out.

In order to correctly understand the scope of the extension, it is essential to make a somewhat more detailed study of its two characteristic and differentiating aspects in regard to art. 3:

- The **limitation of valid reproduction**, outside of which the exception does not function.
- The exception being subject to the **absence of reserved rights**.

Reproduction is not allowed to store the works

“Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining”, says art. 4.2 of the DSMD.

Both literally and strictly interpreted as to making an exception to IP rights, the acts of reproduction allowed are exclusively those necessary to do mining as such (which corresponds to its legal definition), not others. The purposes to which the regulation refers are not the intentions of the person using TDM or the purposes the mining is used for, but exclusively those of mining as such, in its legal conception: analysis.

This interpretation is also supported by the logical-systematic aspect that constitutes its different wording regarding art. 3, a result of the legislator’s differing assessment of the interests at stake: while, in the exception for scientific research purposes, the reproductions and extractions made may be stored for these purposes, with the exception in art. 4 there is no possibility to store the mining for whatever purpose the mining is used for²⁴. That is why, unlike the other, in this point the regulation does not even regulate how this storage should be done. Simply, because it should not be done.

Or, in other words: the **only reproduction allowed under this exception is purely functional and temporary, insofar as it is technically necessary**²⁵, so that text and data mining can produce the patterns, correlations, etc. for which it is used. Such a copy, which is part of the technological process (a necessary step for mining to be done), has no greater economic significance and is, by definition, transient: once analysed, it cannot be retained even for the purposes (commercial or not) for which it was mined.

Sometimes, it is possible to determine that the unauthorised copy has been made (that the model has “retained” the copy of the work) from the system’s own outputs: **If a Gen AI model can reproduce the original work** (or recognizable fragments of it) as a result of its operation, it is because it has stored it (saved, retained) beyond pure technical necessity, against the wording of art. 4 of the

24 In addition to the literal wording of art. 4.3, recital 18, in explaining it, says that copies made should be allowed “to be retained for as long as is necessary for text and data mining purposes” (emphasis our own). By comparison, art. 3 does not speak of retention for mining or mining purposes, but rather of fulfilling the purpose that justifies use of the mining technique, which is broader, in the framework in which it takes place: scientific research. Likewise, recital 15, in relation with retention of copies made for scientific research, when specifically stated that the use of copies for purposes beyond this falls under the general system (where applicable, the exception of art. 5.3.a of the Directive on Copyright in the Information Society, of illustration for educational purposes or scientific research, which could only cover them very residually).

25 By analogy with another more known and regulated for decades: even considering it a reproduction for the purposes of art. 18, what is provisional, without independent economic significance, transitory or accessory and an integral part of the technological process of transmission on the network, does not require authorisation from the rightsholder (art. 31.1 TRLPI).

DSMD and, therefore, ***we are not facing merely an act of analysis to extract metadata or trends*** (object of TDM), ***but rather facing the performance of an act of authentic “memorisation” that goes beyond the scope of analytics*** to enter into the exercise of non-expected exclusive rights: unauthorised reproduction (with the addition of a possible and subsequent public communication, which is also not accepted).

This is what happened, in the court’s judgment, in the case of ***GEMA vs. OpenAI***²⁶: The ruling refused to apply the TDM exception alleged by OpenAI facing GEMA’s claims, arguing that the model’s ability to return fragments of protected works as output shows that they have been used for its training, and that *the TDM exception does not apply to generative AI systems that memorise and reproduce protected content*. For the Court, the TDM techniques accepted are only those that allow the system to reach inferences, patterns, etc., so that everything that goes beyond this limit is not covered by it and obliges users of works or repertoires (such as OpenAI or any other AI developer) to obtain prior authorisation to be able to carry out the training work of such models. This is a strict application of the wording of the regulation (as corresponds to an exception) which, in practice, if maintained, will mean the illegality of many datasets and Gen AI models²⁷.

The distinction is fundamental for Gen AI: ***the Directive legalises analytical reading, not generative reproduction*** (new generation of the original as system output)).

The Gen AI that “creates” from scratch based on what has been learned would act legitimately, even if it had been provided during training or during use of protected works and/or subject matter, provided that in this case it had been done under the exception.

The Gen AI that, in this apparently creative process, “reproduces” (generates again) works and/or subject matter protected by IP, would not do so, because by giving these reproductions as an output (which may also be subject to subsequent unauthorised public communications), it ***would not be covered by the exception, returning us to the general IP rights system*** (which would require prior licensing for the use/exploitation of this type of protected content).

26 Higher Regional Court of Munich (*Landgericht München*), [Judgment of 11 November 2025](#) (*GEMA vs. OpenAI*, Case No. 42 O 14139/24); [Press Release](#), in English. The Court’s position was already put forth in the [initial presentation of the case made in the chamber before the oral proceedings](#).

GEMA (*Gesellschaft für musikalische Aufführungs und mechanische Vervielfältigungsrechte*), filed complaints against **Open AI** (13 November 2024) and **Suno** (21 January 2025, before the same Court), after confirming that the content generated by their respective models or AI agents through simple instructions from their users violated the rights of their associates, as could be seen by simply verifying the outputs, which reproduced protected musical compositions. In the ruling in the case against OpenAI, the Court ruled in favour of the plaintiff, ordering OpenAI to cease unauthorised reproduction and to pay damages, as well as, if it wishes to continue its activity, to obtain licenses for future training.

The *case against Suno* is pending judgment. In this, offered as proof are the evident similarities in harmony, rhythm and melody of world-famous works such as *Forever Young*, by Alphaville, *Mambo No. 5*, by Lou Bega, or *Cheri Cheri Lady*, by Modern Talking, among many others.

27 In the case of *Kneschke vs. LAION*, conversely, the High Court (on appeal) considered a circumstance which, in the specific case, would lead to a different answer: *LAION’s dataset did not store the work, but rather a simple link to it*. That is, instead of retaining it for use, it gives its Internet address so that it can be referred to whenever it is needed. As the dataset only keeps the links, reproduction of the image made, in application of the “rule of three steps” (pp. 21 et seq. of the aforementioned judgment), is considered a simple intermediate technical step, temporary, since it does not affect normal exploitation of the work. What is then done by a Gen AI using this dataset is merely speculative. As such it seems to work like an intermediary services provider of the Information Society, although if so, perhaps the creator of a dataset that works like this, or of the Gen AI that uses it, could be qualified as intermediaries in the sense of art. 17 of the DSMD (providers of online content sharing services that use protected content) and assign them responsibility by this means, by virtue of which it is clear that, if third parties put illegal content online using their services, they are also considered responsible for such acts (under the conditions of the provision).

The Court considers that the opt-out balances the applicable TDM exception, even if in the case it is not really necessary, since it had to be resolved in accordance with art. 3.

The opt-out mechanism and its requirements

“The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online”, establishes art. 4.3 of the DSMD.

Unlike art. 3, the general exception for mining in art. 4, precisely because it is not justified on scientific research purposes by public entities or public interest missions and provides protection for possible commercial uses, may be subject to exclusion by the rightsholders.

The exception only applies if the rightsholder has not expressly reserved the use of their works and/or subject matter, excluding text and data mining done by third parties without their authorisation. This is what is known as the **“opt-out mechanism”**, under which rightsholders can exclude their works from the exception (but if they do not, mining would be allowed²⁸). For this reservation to be valid, the Directive requires that it be:

- **Expressly:** it is up to the rightsholder to show their reservation. It is the *rightsholder who must actively prohibit*, not the AI that must ask for permission.
- **Appropriate:** in principle, according to recital 18, the rightsholder could make their reservation by contract, by unilateral declarations, or in the special form provided for online content (see the following). In any event, concludes recital 18 in explanation of art. 4.3, the rightsholders may establish sufficient measures to guarantee their reservation of text and data mining, but without such measures affecting other lawful uses
- Done by **machine-readable means** when dealing with content accessible online. Given that crawlers scouring millions of websites per second cannot stop to read “Terms and Conditions” in PDF or in a footnote written in natural language, the legislator specifies what it means by “appropriate” manner in the case of online content, requiring that the reservation be expressed in machine-readable means.

The *de facto* historical standard for making such a reservation could be that of the well-known **“no robots”** files (**robots.txt**). It happens, however, that *it is not a protocol designed for IP rights, but rather to avoid search engine indexing*, so its purpose is different and does not allow making distinctions between permitted uses and those that are not. There may be doubts as to whether a “disallow” in robots.txt constitutes (mainstream doctrine) or not (AI companies that deliberately ignore it) an “express” statement for the purposes of this regulation, given its generality. That is why the W3C and other organisations work on more granular standards (e.g. “you can index for searching but not for AI training”, using TDM Rep / C2PA protocols).

Regarding the manner in which the reservation is expressed, there are varied judicial pronouncements, even within the same case. On appeal, and without the need to get into the subject, the High Court of the **Kneschke v. LAION** case made some considerations in regard to this that are very restrictive regarding the way the rightsholders make the reservation: In their view, opt-out should be exercised in a way that it can be functionally implemented in the automated flow of mining, so the use of (natural) human language may not be sufficient. In regard to the date when the events took place (2021, when LLMs had not yet become generalised), the Court notes that natural language was not readable for AI. But even today the same can happen, given the interpretability of human language, which prevents *de facto* the automation of exclusion. Therefore, the judge considers that what is

²⁸ *The case of GEMA vs OpenAI*, p. 14, in relation to the rightsholder: “Bliebe er untätig, sei das Text und Data Mining nach § 44 UrhG erlaubt” (if it remained inactive, TDM would be permitted under this exception).

“appropriate” are the standardised, verifiable and documented mechanisms, which are integrated into the machine’s automatic operational flow when performing mining (metadata and/or universal or extended protocols, such as the “robots.txt” file, which he specifically cites)²⁹.

The understanding of the German Court in this case may be excessive. Certainly, it could apply to online content provided in open form or in repositories where authors can, for their respective works, freely decide (and, therefore, make use of the technical capabilities, labels, etc. of the repository, or those provided by themselves, to make reservation of their rights machine-readable). It seems unacceptable to apply this to online content located behind payment walls, subscriptions or registration, since in this case there indeed must be prior acceptance of contractual or licensing conditions that should require human intervention to be contracted and, thus, there would be effective human knowledge of the limitations and reserves of use made, notwithstanding the fact that they could also have machine-readable means for each work or subject matter. And this, notwithstanding considerations such as those expressed in the initial trial court (if AI is able to “understand” and interpret text, it should also be considered that it understands conditions of access to a website expressed in natural language) or the limitations of current no robots protocols and metadata.

Conversely, the Court’s interpretation of the **GEMA vs. OpenAI case** is more flexible and *pro-rightsholders*, while also addressing the valid possibility of all the associates to reserve rights due to their membership in GEMA. The judgment admits as a fact that the plaintiff stated, after entry into force of 44 UrhG (transposition of art. 4 of the DSMD), a reservation of rights for text and data mining in regard to the works in its repertoire, which has been included since 2024 in the legal notice (*Impressum*) of its website³⁰, and is public, visible and general; additionally, it makes a very clear criticism of the inadequacy of the robots.txt protocols generally used³¹. However, the case ruled in favour of GEMA (*ratio decidendi*) for other reasons (the existence of authentic memorisation of the works, reproduced in the outputs) different from the reservation made, so it does not assess whether the manner in which it is presumed to appear is valid or not.

Lack of compensation for the exception

Nor in this case is there any provision whatsoever in regard to state implementation of any kind of mechanism for the exception, nor any explanation of this. In fact, its absence results, simply, from granting rightsholders the right to reserve and, therefore, of the possibility to obtain their remuneration, if they so wish, through contractual means.

29 Literally, on p. 19 of the judgment: “as for the question of machine readability, it does not only matter that text can be captured by a machine, but that it can also be interpreted by it to the extent that, through an automated procedure, it can be ensured that the contents covered by the reservation are not processed”.

In contrast to this interpretation, the *initial trial court* interpretation, as *obiter dicta*, had been more flexible and favourable to the authors (paragraphs 100-107). Recognising the need for machine readability, it considered that the requirement had been met when the reservation of use is written in “natural language”, believing that there was evidence that even in 2021 there were systems (in particular, for the defendant) to properly read and understand a reservation made in such a way. Basically, the court held that since modern AI technology is capable of understanding complex texts, it should also be considered capable of understanding a prohibition phrase in the terms of use.

30 P. 3. Previously, in May 2022, it had obtained from its associates the power to declare an opt-out to mining, by modifying their management contracts (p. 4), so the court admits that it has the right to do so on behalf of the authors and publishers. Exercise of the reservation took place in the *Impressum* of the GEMA website, that is, in its section where mandatory legal notices are established, making it the most legally visible and formal place on its website, where any user or crawler should be able to find it.

31 “The robots.txt protocol is not fully adequate for declaring reservations of use in the field of text and data mining. With robots.txt it is not possible to make reservations of use referring to specific works, since in this protocol only directories, subdirectories or URLs with certain strings of characters can be generally excluded, but not accurately identify specific content or works. Nor is it possible to deal with types of use or licensing conditions in robots.txt” (p. 14).

Preliminary ruling in progress: *Like Company vs. Google*

In the case of ***Like Company vs. Google Ireland Limited***, the plaintiff accuses Google (Alphabet) of reusing journalistic content without authorisation on its Gemini chatbot, alleging the corresponding copyright infringement and violation of the right of the rightsholders to exclude their works from AI training.

The District Court of Budapest (*Budapest Környéki Törvényszék*) in Hungary referred a [question for a preliminary ruling](#) to the CJEU on 3 April 2025 ([Case C-250/25](#)), asking whether this falls within art. 15 of the DSMD and, in connection with this, (1) if the process of training the chatbot constitutes a real “reproduction” for the purposes of said article and art. 2 of the Directive on Copyright in the Information Society, and (2) whether or not such reproduction of legitimately accessible works does or does not fall within the exception of art. 4 of the DSMD³².

The explanation of the facts in support of the preliminary questions presented explains how this specific chatbot works (paragraph 5, which seems to accept the defendant’s explanation in this regard; see also paragraph 14): “Gemini (Bard) is a basic type of LLM model. It is not an information database nor an information search system. It does not store copies of the collected data, but rather converts the data into tokens—that is, it breaks down the texts into minimal units—and integrates them. This chatbot does not have a fixed database from which it can retrieve any content of the data at the request of users. It makes use of the Google Search database to collect data and also usually suggests that the user search for the topic in question later in Google Search. When asked directly, the chatbot operated by the defendant is able to provide a response that shows the content of a protected press publication”.

According to the plaintiff, “the chatbot used works by Hungarian authors and, during its training, the defendant violated the plaintiff’s reproduction rights, as the quantity and relevance of the data it provided to the program exceeded the limits of the exception” (paragraph 17 of the summary of the request for a preliminary ruling).

The defendant denies that the visualisation of the responses is a reproduction nor made available to the public and that, even if it were a reproduction (paragraph 21 of the summary) “the alleged reproduction would still fall within the exception regulation for acts of provisional reproduction (article 5(1) of Directive 2001/29) and for text and data mining (article 4 of Directive 2019/790). Therefore, the extracts that appear in the reply do not exceed the limits of content and length nor

32 Specifically, the preliminary ruling questions are the following:

1) Should Article 15(1) of Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 [on copyright and related to copyright law and related rights in the digital single market and amending Directives 96/9/EC and 2001/29/EC] and article 3(2) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on harmonisation of certain aspects of copyright and related copyright law rights in the information society be interpreted in the sense that it constitutes a communication to the public the viewing, in the responses of a chatbot based on LLM, of a text partially identical to the content of the websites of press publishers to such an extent that it is already protected under article 15 of Directive 2019/790? If the answer is yes, is it relevant that this is the result of a process in which the chatbot only predicts the next word based on observed patterns?

2) ***Should article 15(1) of Directive 2019/790 and article 2 of Directive 2001/29 be interpreted in the sense that the training process of an LLM-based chatbot constitutes a reproduction that is constructed based on observation and pattern concordance, which makes it possible for the model to learn to recognize linguistic patterns?***

3) ***If the second preliminary ruling question is answered in the affirmative, does such reproduction of legitimately accessible works fall within the exception provided for in article 4 of Directive 2019/790, which guarantees free use for text and data mining purposes?***

4) Should article 15(1) of Directive 2019/790 and article 2 of Directive 2001/29 be interpreted as constituting a reproduction by the chatbot service provider of the fact that, when users give an LLM-based chatbot an instruction that matches or references the text contained in a press publication, this chatbot then generates its response based on the instruction given by the user, a response in which the content of a press publication is fully or partially viewed?

the regulations of exception to legal protection”.

The CJEU's responses to the questions posed by the Court will be decisive for interpretation of the TDM exception and, ultimately, for the relationship between AI developers and rightsholders in Europe.

Comparison Table: art. 3 vs art. 4 of the DSMD

	Article 3 - TDM for scientific research	Article 4 – TDM for general purposes
Legal nature	Mandatory exception and non-waivable by contract.	Mandatory exception, but conditional on an opt-out by rightsholders.
Political objective	Facilitate open science, reproducible research and European academic competitiveness.	Promote commercial innovation, AI and the data economy, balancing interests of the technological and cultural industries.
Regulatory logic	Protect the public interest and remove unjustified barriers to research.	Allow generalised TDM without depriving rightsholders of control over commercial uses.
Legitimising purpose	Scientific research. Strict and limited purpose.	Any purpose, including commercial ones. There is no final limitation.
Beneficiaries	Only research organisations and cultural heritage institutions that are non-profit or have public interest missions.	Any user with lawful access: individuals, companies, startups, Big Tech, public institutions, etc.
Access condition	Lawful access to works (subscription, institutional license, open access).	Lawful access to the works. (subscription, institutional license, open access).
Acts allowed	Reproductions and extractions required for TDM.	Reproductions and extractions required for TDM.
Retention of copies	Permitted for purposes of verification of scientific results.	There is no specific right of retention.
Possibility of opt-out by the rightsholders	No. Rightsholders cannot exclude their works from this exception.	Yes. Rightsholders can reserve rights, expressly, appropriately and by machine-readable means.
Validity of contractual clauses	Contractual clauses cannot cancel the exception.	Contracts can limit TDM if the rights are reserved.
European jurisprudence	LAION case (first instance and appeal)	- LAION case (obiter dicta; first instance and appeal) - GEMA vs OpenAI case (first instance) - GEMA vs Suno case (awaiting judgment from first instance) - Preliminary ruling questions before the CJEU in the case of Like Company vs Google

DIFFERENCES WITH THE U.S. APPROACH (FAIR USE AND COPYRIGHT MARKET DILUTION)

Unlike our IP system, based on recognising certain rights and exceptions, the Anglo-Saxon system allows unauthorised uses of protected works when they meet criteria of public interest and proportionality, known as “fair use”. This is a form of defence against a copyright owner’s claim, based on Section 107 of the U.S. Code (Title 17, Chapter 1), which establishes **four factors to be considered: purpose and character of the use, nature of the copyrighted work, amount and substantiality of the portion used in relation to the copyrighted work as a whole, and effect of the use upon the potential market for or value of the copyrighted work**. Taking these aspects into account, the courts apply a flexible test, on a case-by-case basis, allowing the use of protected works when, specifically:

- It has a transformative purpose, such as criticism, comment, research, or creation of something with a different meaning.
- It does not replace the market of the original work, nor does it significantly affect the value of it.

In the context of Gen AI, this analysis has become essential in determining whether training models with protected works is legitimate or not. In contrast to the claim by model developers of the transformative nature of their technology (the new use contributes something qualitatively different), rightsholders argue on the basis of what is known as “copyright market dilution”: even if the use of their works by AI developers were considered sufficiently transformative, it results in works or subject matter that compete or replace the originals in the relevant markets (exploitation of literary and journalistic works, musicals, audiovisuals, etc.).

In ***Bartz et al. vs. Anthropic*** (Case No 3:24-cv-05417), Judge Alsup appears to accept the position of the defendant, developer of the LLM Claude, when in his [June Order of 2025](#) he admits that there has been fair use, since this technology, which generates new content, is among the most transformative existing (“[t]he technology at issue was among the most transformative many of us will see in our lifetimes”, paragraph 5). The truth is that the case presents some characteristics that do not occur in others, such as the fact that Anthropic acquired thousands of physical copies that it digitalised (lawful access), although it was also alleged by the plaintiffs that Anthropic had also acquired many after digitalising and by using some well-known “pirate” download sites (LibGen and PiLiMi).

Nevertheless, despite the apparent victory for the AI developers, *the judge left the door open to sue the company for piracy*, accepting the facts presented by the plaintiffs about the use of works from datasets having dubious legality and the fact of copying the works themselves: “Anthropic assembled these copies into a central library of its own, copied further various sets and subsets of those library copies to include in various “data mixes,” and used these mixes to train various LLMs. Anthropic kept the library copies in place as a permanent, general-purpose resource even after deciding it would not use certain copies to train LLMs or would never use them again to do so.

All of Anthropic's copying was without plaintiffs' authorization" (Statement and Conclusion). "We will have a trial on the pirated copies used to create Anthropic's central library and the resulting damages, actual or statutory (including for willfulness). That Anthropic later bought a copy of a book it earlier stole off the internet will not absolve it of liability for the theft but it may affect the extent of statutory damages" (Conclusion).

That is, although the use for training was lawful, the download of pirated copies to create the dataset was not.³³

Facing the eventual filing of this class action suit, Anthropic finally reached an agreement with the plaintiffs.³⁴

Conversely, in **Kadrey et al. vs. Meta**,³⁵ Judge Chhabria, of the same District Court and on the same date, made somewhat different considerations in his [order denying the plaintiffs' motion for partial summary judgement](#). Chhabria jointly analysed the download and use of the works to train Meta's LLM model (Llama), considering that there was fair use³⁶, but later accepted the theory of what is known as "*copyright market dilution*": these models allow the creation of thousands of works that, although they do not copy the originals, compete with them³⁷. The problem was that the plaintiffs had failed to provide sufficient evidence of such competition and the damage caused, and thus he could not rule in their favour in the case³⁸.

The market dilution theory "attacks" Gen AI models not from their actions in regard to the exclusive rights of rightsholders (whether the rights are infringed or not), but rather from the parasitic competition such models allow, that without even having obtained permits or licenses for use of the works, they explode into these markets thanks to the prior use done for training³⁹. The U.S.

33 He maintained this analysis in a [denial of a subsequent motion for stay of the case](#) requested by Anthropic (August 2025).

34 [Class Action Settlement Agreement](#) by which it agreed to pay 1.5 billion dollars. After the initial doubts the agreement created for the Judge ([Order No. C 24-05417 WHA, of 7 September](#)), Alsup finally approved it on a preliminary basis ([Memorandum of preliminarily approved settlement](#)) after the parties answered his questions on 25 September 2025. Despite criticism, the CEO of the Association of American Publishers [highlighted the importance of the agreement](#): "*The proposed settlement provides enormous value in sending the message that artificial intelligence companies cannot unlawfully acquire literary works from shadow libraries or other pirate sources to use as the building blocks for their businesses*".

35 Case no. 23-cv-03417-VC, <https://chatgptiseatingtheworld.com/wp-content/uploads/2025/06/Judge-Chhabria-Fair-Use-decision-in-Kadrey-v.-Meta-June-25-2025.pdf>

36 He recognized a light but sufficient public benefit by allowing that the Llama model generates new creative expressions, finding a highly "transformative" use against which even commercial use objections are less relevant: the model does not reproduce the original works (it does not "collect" or "repackage" them), but rather uses them to learn, develop statistical patterns, and then generate new content.

37 "This case, unlike any of those cases, involves a technology that can generate literally millions of secondary works, with a minuscule fraction of the time and creativity used to create the original works it was trained on. No other use—whether it's the creation of a single secondary work or the creation of other digital tools—has anything near the potential to flood the market with competing works the way that LLM training does. And so the concept of market dilution becomes highly relevant (...) it seems likely that market dilution will often cause plaintiffs to decisively win the fourth factor—and thus win the fair use question overall—in cases like this." (p. 32).

38 "The question, then, is whether these particular thirteen plaintiffs in this specific case have presented enough evidence to win on this factor. Or, to put it more precisely given the procedural posture of this case, whether these plaintiffs have presented enough evidence to raise a genuine dispute of material fact sufficient to give the question of market dilution to a jury. The answer is no" (pp. 32-33).

39 The U.S. RIAA (Recording Industry Association of America) itself previously [coordinated several actions and lawsuits](#) against Suno and Udio as early as 2024 with these arguments, for "*massive infringement of sound recordings protected by copyright being copied and exploited without permission by two multi-million dollar music generation services*": **UMG, Capitol, Atlantic et al. vs. Suno** ([lawsuit](#)) filed on 24 June 2024 in the Massachusetts District Court, Case no. 1:24-cv-11611) and **UMG, Capitol, Atlantic et al. vs. Udio** ([lawsuit](#)) with the same date and general content, before the SDNY, Case no. 1:24-cv-04777). For plaintiffs,

Copyright Office itself, in its [Copyright and Artificial Intelligence report, Part 3: Generative AI Training](#) (May 2025) seems to be in favour of embracing the copyright market dilution theory based on the effect of Gen AI on the potential market or value of the protected works: *“The speed and scale at which AI systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data. That means more competition for sales of an author’s works and more difficulty for audiences in finding them. If thousands of AI-generated romance novels are put on the market, fewer of the human-authored romance novels that the AI was trained on are likely to be sold. Royalty pools can also be diluted (...) Market harm can also stem from AI models’ generation of material stylistically similar to works in their training data”* (p. 65).

Regarding the artists, there are ongoing lawsuits, specifically in June 2025 [two other class action lawsuits](#) were filed by country singer Tony Justice and others, [against Suno \(Justice and 5th Wheel Records vs. Suno](#), Case No. 1:25-cv-11739, in Boston, filed 14 June) and [against Udio \(Justice and 5th Wheel Records vs. Udio](#) Case No. 1:25-cv-5026, in New York, SDNY, 15 June) that attack the defendants’ alleged fair use built on the most classic transformative use, counter arguing about competition and market dilution: thus, in Justice’s lawsuit it is argued that *“the fair use doctrine does not excuse unauthorized training on expressive works (e.g., music) particularly when those works are used to generate substitutional outputs that may replace the originals in the relevant marketplace (...) Suno/Udio training its AI model on the copyrighted songs of independent artists, Plaintiffs and Class Members, without authorization, to then create competing music in the exact same marketplace, is not likely fair use, and therefore, is prima facie copyright infringement.”*

Meanwhile, the first agreements between AI developers and the cultural industry are beginning to take place in the U.S., in particular, associations and music labels (majors as well as indies). There are large rightsholders who have licensed their own content, such as Disney ([with OpenAI](#), for their Sora model), Shutterstock (with [Meta](#), [OpenAI](#) and [NVIDIA](#), to use their catalogue to train and improve their respective models), Getty Images (with [Perplexity](#), a global agreement to display images from their catalogue in their AI-powered search results); there are press groups that have also reached agreements, under which models from OpenAI⁴⁰ and Meta⁴¹ can feed their Gen AI systems and use their content for training and to display fragments, quotes and links; several majors (UMG, Warner, Sony...) have signed agreements with developers such as Udio (an AI-based music creation platform, as such putting an end to the lawsuits they had against it)⁴² and with YouTube/

the alleged fair use does not apply to Gen AI because it produces results that directly compete with the copyrighted materials that have been used to train it. Fair use, [affirms the RIAA](#), *“is not available when the output seeks to ‘substitute’ for the work copied. And Suno and Udio have, in their own words, conceded that is exactly what they intend,”* the RIAA asserted”. In addition, they note, AI developers and investors themselves have admitted that they use protected content to train their models. ECSA (the European Composer and Songwriter Alliance) [welcomed](#) the filing of these lawsuits.

Udio has been reaching a number of agreements with the plaintiffs as a way to end these lawsuits.

40 Associated Press, News Corp, Financial Times, The Atlantic Vox Media, Axel Springer... A chart of the agreements up to June 2024 can be found on the [Mx Media Alliance website](#). OpenAI has an [agreement with Spanish company Prisa Media](#) as of March 2024, under which ChatGPT users can interact with content from the Group, and these users are also useful to train and improve the model.

41 See [TechCrunch](#).

42 Under these agreements, Udio will pay for a commercial license to access the content from the catalogues of these majors to create a new platform to exploit AI features. As has been reported (these agreements are usually kept private between the parties), these pacts include an opt-in system (authors and artists must also consent for their tracks to be accessed), the labelling of results created with AI (fingerprinting) and consumption (downloading) only on the platform (what is known as a “walled garden”). Reactions to such agreements are mixed: Udio sells them as a commitment to the music industry, which also considers them a success and a way to monetize Gen AI, while some [users](#) criticize them for involving, they say, the [destruction](#) of the Udio model, which would become a type of streaming; nor are they to the liking of some IP experts and rightsholders (a system of

Google, through the “Music AI Incubator” program, to experiment with AI models that use protected musical catalogues in a controlled and remunerated framework⁴³.

In regard to independent music, Merlin (an organization representing thousands of independent labels and distributors around the world) has signed an [agreement](#) with Udio to license recordings by Merlin members who choose to participate in training AI models, under an opt-in model (each label or independent artist chooses whether it wants its catalogue to be used, with financial compensation for those involved, protections against misuse, and to preserve the value of human authorship and the intention to explore new forms of musical expression generated with AI⁴⁴.

illegal use is “laundered” and the end of the trial has not been reached to specify the concept and requirements of fair use), while there are [musicians who fear](#) that the artists on these labels will be favoured over everyone else.

43 The initiative was built on the partnership between YouTube and UMG. According to its [official blog](#), the program is governed by three fundamental principles published by the YouTube Blog: embracing AI responsibly, including protections for artists, and maintaining robust security and trust policies. YouTube has had conversations to incorporate Sony Music and Warner Music Group.

44 Despite the fact that this is the agreement that has gotten the most attention, given the importance of the counterpart (as Udio is), Merlin had already reached another agreement with ElevenLabs, under the same opt-in premise (see news articles on this from [Billboard](#) and [MBW](#)).

RECOMMENDATIONS ON THE TDM EXCEPTION AND ITS EVOLUTION IN GENERATIVE AI

1. Review the regulatory model of the TDM exception

In Europe, cases such as GEMA vs. OpenAI reveal a progressive but firm paradigm shift in assessment of the functioning of Gen AI, which makes the TDM exception insufficient to protect it legally: when the model goes beyond simple “analytical reading” or pattern extraction to retain enough of the work with which it has been trained to supply it in the output, we are already speaking of “memorisation”, copying or unauthorised reproduction. The TDM exception would not even apply. This fact highlights the **foundational defect of the legislative model**: it was born already obsolete or, at least, it is not sufficient to adequately balance the interests of the technological and cultural industries.

The shortcomings of the TDM exception have recently been made clear by the European Parliament, which claims⁴⁵ that current TDM rules were not envisioned for⁴⁶ nor are useful for Gen AI⁴⁷, admitting that there is evidence of generalised violations of the exception itself by Gen AI developers⁴⁸.

2. Establish an effective system for reservation of rights (opt-outs) and guarantee they are respected

Doubts on the way in which reservation of rights should be expressed by rightsholders, made abundantly clear by the disparate judicial interpretations, show the necessity to establish an effective opt-out system, as the European Parliament has also highlighted.⁴⁹

Additionally, rightsholders have serious doubts about the respect of certain Gen AI developers for their reserved rights, added to which is the problem of whether access to content from sets

45 *The Voss Report*, named for its author: European Parliament (Committee on Legal Affairs), *Report A10-0019/2026 on copyright and generative artificial intelligence – opportunities and challenges* (2025/2058(INI)), of 25 February 2026. Available at https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.pdf

46 “The European legislator incorporated Article 4 in the AI Act without, however, having clearly established the consequences. The current exemption in Article 4 allowing text and data mining under the conditions specified **was not drafted with the intention of enabling the use en masse of copyright-protected material by all through generative AI**. And certainly **not where it also leads to the creation of a competitive product accessible to the public**” (*Explanatory Statement*, p. 19).

47 “The current systems for the reservation of rights (‘opt out’) to copyright-protected content are **often impractical**, may not cover all relevant acts of **text and data mining**, and **lack the necessary transparency** for effective implementation and enforcement” (paragraph K, p. 7).

48 As well as other illegal behaviour, such as using pirated sources to get content to train the models: “There is evidence of the **widespread violation of copyright rules** by GenAI providers, including the unauthorised collection of works from the internet, the **non-compliance with rights holders’ text and data mining rights reservations**, the **use of pirated sources to obtain works**, and the failure to seek licences” (paragraph L, p. 7).“

49 The Parliament proposes a standardised system, with a European registry for opt-outs in machine-readable formats, as rightsholders should have tools “to effectively exclude the use of their works in AI training in a limited number of standardised machine-readable formats managed by a trusted intermediary”, such as the European Union Intellectual Property Office (EUIPO; paragraph 10, p. 13).

or datasets on which suspicions of illegality are more than well-founded is lawful. If these are confirmed, it should not be so, even if such sets were free access, which reinforces the need for effective control and enforcement mechanisms.

3. Strengthen transparency obligations in the use of content for TDM

Likewise, effective protection for rightsholders requires demanding total transparency from those who are providing and implementing the AI models and systems in regard to the content used in text and data mining⁵⁰. This obligation could be reinforced by means of a presumption *iuris tantum* on use of protected works and subject matter when there is not transparency, as proposed by the European Parliament⁵¹.

4. Evolve from the opt-out system toward ex ante (opt-in) authorisation mechanisms

If interpretation of the Munich Court is consolidated, the “opt-out” system (where the author must hide from the AI) could be transitioning via a judicial path toward a *de facto* “opt-in” requirement for generative models capable of reproducing literary or musical works, forcing negotiation of licenses *ex-ante*. But it is still a jurisprudential doctrine that should receive legislative support, with a change in the regulatory model.

5. Establish compensation mechanisms for past uses

Furthermore, any reservation of rights on the use of TDM has only been possible, over time, since the Directive came into force. And as well only considering the general uses of data mining that this regulation contemplated, which, as shown, did not include Gen AI. This has made use of the mining exception instrumental, in regard to the demands on it from what it does. That is to say, the reserving of rights in the face of such actions is very recent and the possibility of doing so was not even known by many authors and/or artists, that is, small non-industrial rightsholders. But even big publishers and music industry majors were surprised by the technology, such that many datasets and Gen AI models were trained *before* the rightsholders could implement the opt-outs. In such cases, future legislative solutions or trade agreements may provide little or no compensation to rightsholders whose rights have *already* been damaged, which makes it necessary to explore specific compensation mechanisms.

6. Develop a new licensing framework for generative AI

On the other hand, in a context of increasing litigation, ***the signing of agreements*** between Gen AI developers and rightsholders of protected content of all kinds ***highlights the growing awareness that indeed use is being made of reserved rights exclusively by rightsholders***, so it is therefore necessary to construct a new licensing framework⁵².

⁵⁰ In accordance with the Voss Report, «transparency concerning all copyright-protected content used to train those models and systems (...) must consist of an itemised list identifying each item of copyright-protected content used for training (...) AI providers or deployers are required to maintain detailed records of their crawling activities» (paragraph Y, p. 9).

⁵¹ In absence of the understood transparency as described (itemised list of each item used), it should be presumed (*iuris tantum*, with possibility for rebuttal) that indeed protected works have been used to train their model: not acting transparently “should give rise to a **rebuttable presumption** that any relevant copyright-protected work... **has been used** for the purposes of training, inferencing or retrieval-augmented generation” (paragraph AB, p. 10). In this sense there are also advances in France: see analysis of their legislative proposal regarding this, currently in the legislative process: <https://genaie.es/ley-francesa-so-bre-ia-derechos-autor-presuncion-uso-obras-protegidas/>

⁵² Being the current copyright law framework “**insufficient to address the challenge of licensing copyrighted material for GenAI**”, the Parliament “calls for an **additional legal framework** to clarify licensing rules”, and “insists that such a framework

7. Evaluate extended collective licensing mechanisms

In this regard, and in view of the unequal negotiating position of the parties, other compulsory licensing mechanisms to be managed by collective rights management bodies are also being studied, such as “**extended collective licenses**”, that would allow management entities to grant non-exclusive authorisations to AI developers for the use of works protected by copyright law, even if the rightsholders had not given an express mandate to the entity⁵³.

8. Establish measures against models trained without respecting rights

Furthermore, and in line with the positions of the European Parliament, the possibility of excluding models trained without respecting copyright law from the European market⁵⁴ has been raised, thus strengthening effectiveness of the system and preventing that external agents benefit from regulations with a laxer protection of rights to compete with an unfair advantage over the European models or models that fully respect our regulations.

9. Extend the protection for performers beyond classical intellectual property

In regard to performers, ***the defence of their rights and interests via Intellectual Property may not be sufficient, at least in its current configuration***. Their services engage their voice and image, which must also be protected because their unauthorised use by third parties may not only damage their rights over their interpretations (reproduction, public communication, transformation...), but also reputational damage (to their image rights and honour) and even a threat to their work. Thus, like what happens with the works, the content generated using them can end up competing in the market with their original performances and even replace them.

Notwithstanding ***updates to labour regulations***⁵⁵ ***and image rights***⁵⁶, and the fight against so-called “deepfakes” in the AI regulations themselves (arts. 3(60) and 50.4 RIA), ***in some regulations it is currently being proposed to accept image and voice among the intellectual property rights***

should include provisions ensuring the effective cooperation of GenAI providers with creators and other rightsholders, including **a functioning licensing market that restores the bargaining power of rightsholders and viable protection solutions**” (paragraph 2, p. 11).

53 See art. 12 and recitals 44 and 45 of the DSMD. In Spain, at the beginning of 2025, the Ministry of Culture withdrew the Royal Decree draft law which sought to establish a system of extended collective licenses for the mass exploitation of works and subject matter protected by intellectual property rights for the development of general purpose AI models, faced with the rejection shown by the sector itself.

In the aforementioned Report from the European Parliament, both the possibility of establishing voluntary collective licensing agreements by sector (paragraph 9, p. 13) as well as a presumption of collective management of rights in respect to AI-related uses (which, without being a mandatory collective license, would indeed allow, unless opposed by the rightsholder, that the management entity could negotiate on behalf of the rightsholders; paragraph 6, p. 12).

54 “In order to ensure the proper implementation of EU law in relation to all generative AI services and to prevent non-EU-based providers from gaining an unfair competitive advantage through non-compliance, it is essential that rules protecting and enforcing authors’ rights at EU and national level apply effectively to all services operating on the EU market; (...) GenAI systems that do not comply with these requirements should be barred from operation within the EU and these principles should be robustly enforced” (paragraph AH, p. 11).

“EU copyright law, as recalled in recital 106 of the AI Act, applies regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those GenAI models and systems take place, with the consequence that, where copyright is not observed, those models and systems are barred from being placed or made available on the EU market” (paragraph 16, pp. 14-15).

55 See [GenAle’s analysis of the projected new Artist Statute](#) in the area of AI, contained in art. 13 of the proposed *Royal Decree regulating the special labour relationship of artists who develop their activity in the performing, audiovisual and musical arts, as well as of persons who perform technical or auxiliary activities necessary for the development of such activity*.

56 Art. 8 of the *Preliminary Draft of Organic Law on civil protection of the right to honour, to personal and family privacy and to one’s own image*.

of artists. Although this does not seem to be a viable option at the EU level⁵⁷ and it is outside of the traditional system and concept of intellectual property in Spain, there are already proposals on this in the United States⁵⁸, Denmark⁵⁹ and Mexico⁶⁰ that can serve as a comparative reference and which could be considered should EU or Spanish legislators decide to also follow this path.

In any case, it is worth disseminating these other alternative routes on intellectual property among artists so they can complement it to protect their economic and moral interests and perhaps resolve cases in which copyright and related rights are not enough in their current configuration.

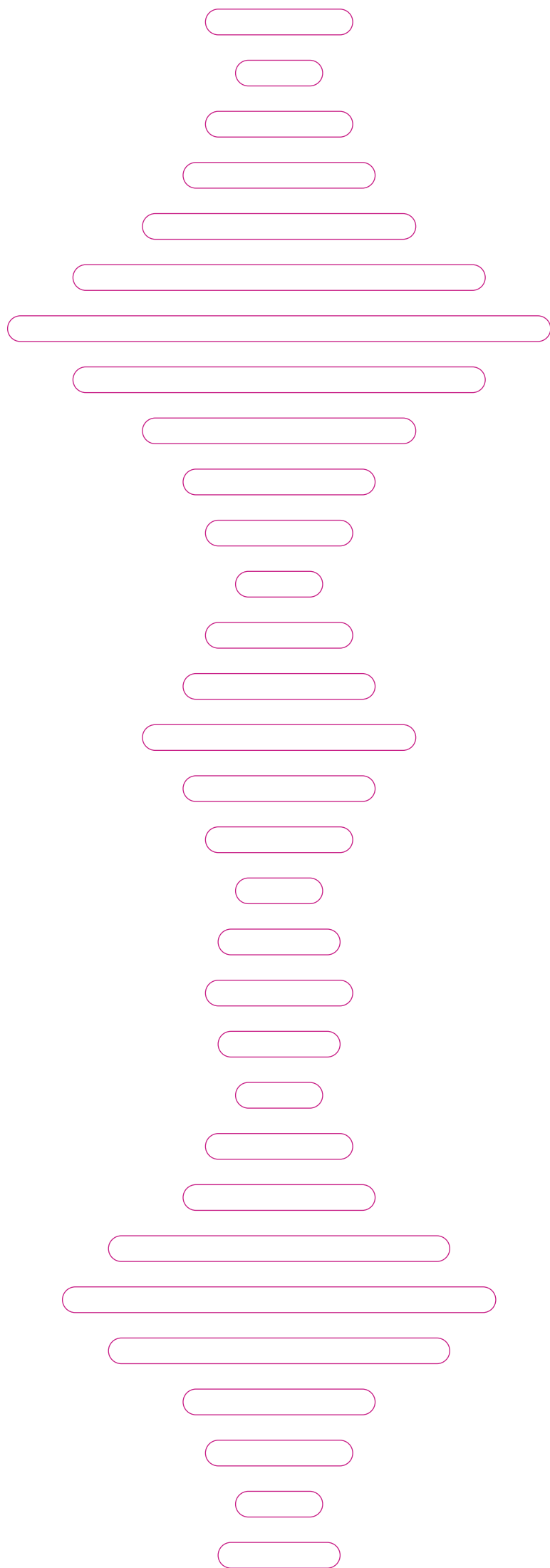
57 According to the [questionnaire](#) drawn up by the Danish EU Presidency, most EU countries are not in favour of using copyright law as a tool to address the risks posed by AI deepfakes for artists and performers; see also the news article in [Euractiv](#).

58 [Draft Law No AI FRAUD Act](#) (*No Artificial Intelligence Fake Replicas and Unauthorized Duplications Act*, H. R. 6943) introduced in the U.S. Congress in January 2024 (previous 118th legislature) with the intention of recognising for every person an “intellectual property right” (sic) over their appearance and voice; it has not been re-introduced into Congress in the current legislature.

Although not exactly covered as intellectual property rights, in Tennessee, the [ELVIS Act](#) (*Ensuring Likeness, Voice, and Image Security Act*, HB 2091 / SB 2096) of 2024, which prohibits unauthorised commercial exploitation of name, image (photography), voice and appearance, over which a real *property right* is recognised, enabling the rightsholder to control their exploitation; specifically, “voice” means sound in any medium that is recognizable and attributable to a person, containing said sound their own voice or a simulation.

59 The Danish government has announced its intention to legislate against deepfakes and protect digital identity. Following various ministerial statements, different media outlets ([AP News](#), [Time](#), [The Guardian](#)...) and specialized *blogs* ([Clovers Law](#)) have published that, among the measures to be adopted, reform of their IP regulations is being considered to include the image, voice and other personal traits within copyright law protection, which is also consistent with the Danish position in Europe, when asked about it.

60 [Advertisements](#) from an initiative to protect dubbing actors and broadcasters made by President Sheinbaum and the Secretary of Culture in February 2026, through a reform of the Federal Labour and Copyright Laws to recognize the human voice as a unique and unrepeatable artistic tool, which is part of the image of the person and whose use requires express and informed authorisation before its use by AI (voice cloning is [prohibited](#) without the express authorisation of its exclusive rightsholders).





www.genaie.es

