UNIVERSITÀ
DI PAVIA

DIPARTIMENTO DI GIURISPRUDENZA
CORSO DI LAUREA MAGISTRALE IN GIURISPRUDENZA

Generative AI Training within EU and Italian Copyright Law: Systematic Analysis through Directive CDSM, AI Act and Law No. 132/2025

Relatore: Chiar.mo Prof. Michele Bertani

Tesi di laurea di
Virginia Perna
Matr. 501761

Anno accademico 2025/2026

INDEX

ABSTRACT

La presente tesi esamina se, e a quali condizioni, l'addestramento di modelli di intelligenza artificiale generativa (G-AI) possa essere ricondotto al quadro normativo vigente in materia di diritto d'autore, tanto a livello europeo quanto italiano, con particolare attenzione all'eccezione per l'estrazione di testo e dati (TDM) introdotta dalla Direttiva (UE) 2019/790 (Direttiva CDSM), alla sua interazione con il Regolamento (UE) 2024/1689 (AI Act) e, con riguardo alla normativa italiana, alla recente Legge n. 132/2025.

Il primo capitolo ricostruisce l'evoluzione del diritto d'autore europeo nell'ambiente digitale e analizza i diritti potenzialmente coinvolti nell'addestramento di modelli di G-AI, con particolare attenzione al diritto di riproduzione e al diritto di estrazione sui generis sulle banche dati. Una questione centrale affrontata nel capitolo riguarda se i modelli di G-AI addestrati costituiscano di per sé una forma di riproduzione — questione su cui la dottrina e la giurisprudenza emergente sono divise. Successivamente il capitolo affronta le eccezioni potenzialmente applicabili alla G-AI, soffermandosi in particolare sull'eccezione di TDM.

Il secondo capitolo presuppone l'applicabilità, almeno in parte, dell'eccezione di TDM all'addestramento di G-AI e affronta le criticità giuridiche e pratiche che emergono da tale estensione. Fra queste viene dedicata particolare attenzione ai problemi relativi all'operatività del meccanismo di opt-out (art. 4(3) CDSM) e all'adeguatezza degli obblighi di trasparenza e di policy introdotti dall'art. 53 AI Act e integrati dal Codice di condotta per i modelli IA per i fornitori di general-purpose IA (GPAI). Inoltre, il capitolo si occupa delle criticità relative all'implementazione della Direttiva CDSM e dell'AI Act, evidenziando come la catena di approvvigionamento dell'IA, prima di arrivare al modello, sia molto lunga e opaca e sia spesso difficile verificare le veridicità delle dichiarazioni dei fornitori di GPAI. Il capitolo si chiude esaminando le proposte legislative avanzate

per sopperire alle difficoltà applicative, con particolare riferimento alla risoluzione del Parlamento europeo del 10 marzo 2026.

Il terzo capitolo si concentra sul contesto normativo italiano alla luce della recente Legge n. 132/2025, che introduce l'art. 70-septies nella Legge n. 633/1941. Tale disposizione è di particolare rilevanza, poichè affronta la relazione fra l'eccezione di TDM e l'addestramento di modelli e sistemi di G-AI. L'analisi si concentra sulla struttura della norma, sul suo iter legislativo e sulle questioni interpretative emergenti, operando un confronto con il quadro europeo.

Il quadro complessivo che emerge da questo studio è quello di un ordinamento che offre una base normativa per ricondurre l'addestramento di G-AI all'eccezione TDM, almeno in parte, e a determinate condizioni, ma che presenta criticità strutturali significative — sul piano del meccanismo di opt-out, della trasparenza e dell'enforcement — che non sono state ancora risolte né a livello europeo né a livello nazionale. I prossimi sviluppi legislativi e giurisprudenziali, in particolare l'attesa pronuncia della Corte di giustizia nel caso Like Company v. Google e la valutazione annunciata dalla Commissione europea a seguito della risoluzione parlamentare del 10 marzo 2026, saranno determinanti per definire se e in quale misura il quadro vigente potrà essere considerato adeguato.

INTRODUCTION

Over the past four years, the rapid emergence of generative AI has had a profound impact on many fields of human knowledge and work, and it has also given rise to significant concerns of a legal, social, and economic nature. As with previous technological innovations, it has introduced important transformations. Nevertheless, unlike earlier revolutions, generative AI seems to affect creativity, a domain that has traditionally been regarded as distinctly human. Indeed, previous waves of automation either replaced physical labour or, in relation to more intellectual activities, affected repetitive tasks. Generative AI, by contrast, is able to produce outputs that closely resemble human creative activity. This is particularly evident in the case of general-purpose AI, that is, under EU law, an AI system capable of performing a broad variety of tasks and interacting with users in a way that simulates natural language.

The key to the capabilities of this technology is the training process underlying it. Generative AI models are developed through the ingestion of vast amounts of data, which are processed and statistically analysed through machine learning techniques or, more typically, through deep learning architectures based on neural networks (transformers). Thanks to this process, models learn recurring structures and patterns in language and are thus able to generate outputs on a probability basis.

This technological development has impacted several areas, including the labour market, privacy, public security, and copyright law. This thesis focuses on the latter, which gives rise to a twofold problem. On the one hand, the datasets used to train generative AI models may include copyright-protected materials, which are copied, cleaned, transformed, processed, and analysed as part of the training pipeline. This first part of the process is not new and was already known in the context of data science. On the other hand, there is a teleological problem: the purpose is not merely analytical: the technology aims at generating new content.

This may compete with the original training material and, in some cases, it may even reproduce it in an exact or near-exact way in the output.

These activities give rise to interpretative challenges within the current copyright legal framework. Indeed, different rights may be implicated in the process, and it becomes necessary to classify these activities within existing copyright law. In this respect, the European Union has, over the years, adopted a number of legislative instruments addressing copyright in the digital environment, providing legal tools that may also be relevant to generative AI training. Among these, particular attention must be paid to the text and data mining exception.

The rationale behind the introduction of the text and data mining exception was to foster and facilitate data science. The latter may be understood as the integration of scientific discovery and practical application through the collection, processing, analysis, and interpretation of large volumes of data, with the aim of extracting information or making predictions. Of course, this explanation belongs more properly to the field of scientific research; however, the exception was also adopted in relation to any other purpose.

The debate on generative AI training and copyright mostly concerns the applicability of the text and data mining exception to such training. Indeed, its legal definition is intentionally very broad and capable of encompassing a wide variety of activities. This debate has been addressed by legal scholarship, case law, and national and regional (EU) legislators. Some argue in favour of a broad application of the exception, whereas others maintain that such an approach is not always justified, either from a technical or from a legal perspective.

It is noteworthy that the European Union has adopted the AI Act, a legislation specifically addressing AI. Even though the latter does not directly regulate copyright, it contains certain provisions that may influence the interpretation of earlier legislation, in particular the TDM exception, especially in light of the related documents.

Moreover, there are responses at a national level that address the intersection between copyright and G-AI training. Indeed, Italy has recently enacted Law No. 132/2025, which introduced Article 70-septies into the Italian Copyright Act — the first explicit national provision addressing the application of the TDM exception to AI training within the EU legal order. This provision raises significant questions regarding its actual consequences and its effectiveness.

It should be noted that the question of G-AI training and copyright is not exclusively a European concern. Other major jurisdictions have adopted different approaches to the same problem: the United States relies on the fair use doctrine, which has been invoked in several pending litigations involving major AI developers, while Japan has introduced a broad TDM exception that has been interpreted as largely permissive of AI training activities. These divergent regulatory choices raise questions about the competitive implications of the European framework and the risk that training activities migrate to more permissive jurisdictions, with potential consequences for the protection of European rights holders. These comparative dimensions, however, fall outside the scope of the present thesis, which focuses exclusively on the EU framework and its Italian transposition.

Against this background, the thesis is structured as follows. The first chapter reconstructs the evolution of EU copyright law in the digital environment and introduces the legal and technical foundations of G-AI training, with particular attention to the role of the text and data mining exception. The second chapter moves to the next stage: assuming that the TDM exception applies, at least in part, it analyses the technical and legal challenges relating to lawful access, opt-out mechanism, transparency obligations, and enforcement thereof. The third chapter analyses the Italian legislative framework for G-AI, particularly in light of recent reform.

In conclusion, the thesis aims to clarify whether, and under what conditions, generative AI training can be accommodated within the existing EU and Italian copyright framework.

CHAPTER I – Generative AI in the EU Copyright Framework

Summary: 1. Historical and Legislative Context; 1.1. The Evolution of EU Copyright in the Digital Environment; 1.2. From the Transistor to the Transformer: G-AI Revolution; 2. G-AI and Its Implications for Copyright; 2.1. Rights Potentially Implicated by G-AI Training; 2.1.1. Reproduction Right; 2.1.2. Extraction Right; 2.2. Assessing the Applicability of Exceptions; 2.2.1. Temporary Acts of Reproduction; 2.2.2. Is TDM Exception Applicable to G-AI Training?; 2.2.2.1. Is G-AI Training TDM?; 2.2.2.2. Structural Limits of the TDM Exception in Relation to G-AI Training

1. Historical and Legislative Context

1.1. The Evolution of EU Copyright in the Digital Environment

Since the advent of the internet, it became increasingly evident that digital technologies were affecting copyright law.[1] In order to keep up with these developments, over the past forty years, the European Union (hereinafter, EU) has intervened multiple times on the topic. The main aim was to harmonise copyright rules across its Member States and to strike a balance between the protection of rightholders and the promotion of innovation. In relation to the former, it became increasingly clear from the Court of Justice (hereinafter, CJEU) jurisprudence[2] that the copyright national legislations were relevant for the internal market. Noticeably, there were concerns about the global competitiveness of the European Communities (now, European Union). This emerged clearly with the *Green Paper on Copyright and the Challenge of Technology— Copyright Issues Requiring Immediate Action* presented by the Commission of the European Communities in 1988.[3]

[1] M. Bertani, A. Cogo, P. Fabbio, A. Genovese, A. Ottolia, Lineamenti di diritto industriale. Concorrenza e proprietà intellettuale, CEDAM, Padova, 2024, Parte VII, cap. V, §1.

[2] CJEU, EMI Electrola GmbH v Patricia Im- und Export Verwaltungsgesellschaft mbH, C-341/87, EU:C:1989:30.

[3] Commission of the European Communities, Green Paper on Copyright and the Challenge of Technology — Copyright Issues Requiring Immediate Action, COM (88) 172 final.

Following the Green Paper of 1988, several directives were adopted during the 1990s to regulate specific subject matters, including software and databases.[4] In 1995, the Commission presented a new Green Paper on Copyright and Related Rights in the Information Society,[5] with which it highlighted the necessity for even deeper harmonization, in particular with regard to the information society.[6] This culminated in the adoption of Directive 2001/29/EC (hereinafter, Infosoc Directive), which introduced a broader degree of harmonisation with regard to economic rights, related rights, and the system of exceptions and limitations. At the same time, during those years the European Communities signed the World Intellectual Property Organization (hereinafter, WIPO) Copyright Treaty (hereinafter, WCT), acknowledging the necessity for an international level of regulation as well (*infra* in this section).

However, after the adoption of the Infosoc Directive, EU copyright harmonisation entered a phase of relative stagnation, with only limited legislative developments for several years.[7] During that period, the internet continued to expand rapidly, while the European Union itself underwent major institutional and economic changes.

Against this background, the debate on copyright reform re-emerged at the end of the decade, with a Reflection Document held by the Commission. The latter draws attention to the necessity to address online users' creativity and deal with

[4] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 27.3.1996, pp. 20–28; Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs, OJ L 122, 17.5.1991, pp. 42–46.

[5] European Commission, Green Paper on Copyright and Related Rights in the Information Society, COM(95) 382 final, 27 July 1995.

[6] The concept of the information society, as employed in EU law, broadly refers to a digital environment in which information and knowledge are produced, processed, and disseminated through electronic means on a large scale.

[7] With the exception of Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights, OJ L 157, 30.4.2004, pp. 45–86. For further analysis, see Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press, p. 3.

dematerialization challenges, and it also contains a reference to a full harmonization of EU copyright laws.[8]

In the following decade there have been major changes; in particular, with the launch of the Digital Single Market Strategy in 2015.[9] Within that strategy, the reform of the EU copyright framework occupied a central position.[10] The rationale behind this initiative was further clarified in the European Commission's *Communication Towards a Modern, More European Copyright Framework* (2015), in which the Commission stated that it was necessary to adapt copyright rules to new technological realities so that they could continue to fulfil their objectives.[11]

Shortly thereafter, the Commission presented the *Proposal for a Directive on Copyright in the Digital Single Market* (2016). The proposal consisted in forty-seven recitals and twenty-four articles, and it was accompanied by an Explanatory Memorandum that again stressed the need to adapt EU copyright rules, including rights, limitations, and exceptions, to the technological landscape of the time.

The legislative process ultimately concluded in 2019 with the adoption of Directive (EU) 2019/790 (hereinafter, CDSM Directive), following extensive amendments during the legislative negotiations. The final text, consisting of eighty-six recitals and thirty-two articles, reflects a move towards a significantly more detailed and articulated legislative framework.[12]

[8] European Commission, Reflection Document of DG INFSO and DG MARKT on Creative Content in a European Digital Single Market: Challenges for the Future,22 October 2009.
[9] Note that as early as 2011, the Commission had published the Single Market Act and the Communication A Single Market for Intellectual Property Rights, COM(2011) 287 final, though these initiatives did not lead to significant legislative change in the field.
[10] European Commission, Communication: Towards a Modern, More European Copyright Framework, COM(2015) 626 final, §3.3.
[11] Ibidem, §1 "[...]it is necessary to: [...] adapt copyright rules to new technological realities so that the rules continue to meet their objectives.
[12] Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press.

In addition to its internal copyright framework, the European Union is also bound by international agreements. Notably, it is a member of the World Trade Organization (WTO), established by the Marrakesh Agreement,[13] and is therefore bound by the agreements annexed to it, including the TRIPS Agreement, which forms Annex 1C.[14][15] Furthermore, Article 9(1) TRIPS obliges WTO Members to comply with Articles 1 to 21 of the Berne Convention (1971) and the Appendix hereto, with the exception of Article 6bis. This is of consequence because, although the EU is not a contracting party to the Berne Convention, it is nonetheless bound by its substantive standards through TRIPS. As previously stated, the EU is also a Contracting Party to the WCT, which specifically addresses copyright in the digital environment, and likewise incorporates the substantive provisions of the Berne Convention.[16][17] This legislation is of particular importance for the purposes of this thesis, as it represents WIPO's response to the challenges raised by digital technologies.

The relevance of these international instruments is evidenced by the observation that the EU regulates only certain aspects of copyright, particularly in the context of the digital environment, and does not provide a comprehensive and uniform body of general copyright rules. In this regard, adherence to these treaties establishes a minimum common standard for the EU copyright landscape, which is also relevant to the interpretative role of the Court of Justice.

[13] World Trade Organization, Marrakesh Agreement Establishing the World Trade Organization, signed at Marrakesh, 15 April 1994, entered into force 1 January 1995.
[14] Council Decision 94/800/EC of 22 December 1994 concerning the conclusion on behalf of the European Community of the agreements reached in the Uruguay Round multilateral negotiations, OJ L 336, 23.12.1994, pp. 1–2.
[15] Agreement on Trade-Related Aspects of Intellectual Property Rights, Annex 1C to the Marrakesh Agreement.
[16] WIPO Copyright Treaty, adopted in Geneva on 20 December 1996; approved on behalf of the European Community by Council Decision 2000/278/EC of 16 March 2000, OJ L 89, 11.4.2000, pp. 6–7.
[17] M. Bertani, Diritto d'autore europeo, Giappichelli, Torino, 2011 (Quaderni di AIDA, 21), cap. I, §5.

An understanding of the rationale and history behind the main EU and international legislation for digital copyright is essential to comprehend the classification and regulation of Generative AI.

1.2. From the Transistor to the Transformer: G-AI Revolution

Following the adoption of the CDSM Directive, technological innovation did not stop. Although the public breakthrough of generative AI is commonly associated with the launch of ChatGPT in November 2022, it is important to note that the underlying technology had been developing for years.

The emergence of generative AI (hereinafter, G-AI) rapidly stimulated public and academic debate across a wide range of fields, from ethics to science, and soon attracted the attention of EU institutions as well. However, it should be clarified that the EU had already begun to reflect on the regulation of AI before the widespread diffusion of G-AI. In particular, the European Commission presented its Proposal for an Artificial Intelligence Act in 2021.[18] Nevertheless, the Proposal was not specifically designed around G-AI, nor did it yet fully address the distinct implications of general-purpose models (hereinafter, GPAI)[19]. Rather, it reflected an understanding of AI as a broader and largely sector-specific phenomenon,[20] whose risks were already emerging in several areas of law and policy. Furthermore, the increasing relevance of AI had already been acknowledged in earlier instruments and international policy documents, including the OECD

[18] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final; see also the accompanying Impact Assessment, SWD(2021) 84 final.
[19] For the purposes of this thesis, it is important to distinguish between AI, G-AI, and GPAI. AI broadly refers to systems designed to analyse data and perform tasks such as prediction, classification, recommendation, or decision-making. G-AI, instead, refers to a specific subset of AI models capable of producing new content on the basis of patterns learned from large-scale datasets. This difference is legally significant, since it affects both the training process and the nature of the output. The notion of GPAI, however, belongs to a different perspective. Under the AI Act, it identifies a regulatory category centred on a model's significant generality and its capacity to perform a wide range of distinct tasks, rather than on its specific technological function. The two notions of G-AI and GPAI therefore do not fully overlap, although in practice many of the most relevant general-purpose models are also generative.
[20] Almada, M. (2025). The EU AI Act in a global perspective. In M. Furendal & A. Lundgren (Eds.), Handbook on the global governance of AI. Edward Elgar (forthcoming). Draft of 17 June 2025.

Recommendation on Artificial Intelligence.[21] It is also noteworthy that EU institutions of that time were aware, even before the rise of G-AI as a mass phenomenon, of the possible impact of AI on copyright. This emerges from the European Parliament Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies.[22] In this context, the Parliament identified the need to strike a balance between the promotion of innovation and the safeguarding of intellectual property rights, which, as we have seen, might be said to be the leitmotif of copyright regulation in the digital age. Although the Resolution is primarily concerned with the protection and governance of AI-related innovation, it also contains significant references to the position of rightholders whose protected works may be used in the development of AI technologies. In particular, it acknowledges the relevance of existing text and data mining exceptions,[23] while also noticing the risk that authors whose works are used to power AI systems may not be fairly remunerated. [24]

At that stage, however, the discussion remained relatively general. The Commission proposal did not contain specific copyright provisions in relation to AI, nor addressing the training of G-AI or GPAI models, as stressed above. [25] Copyright concerns were only indirectly reflected in the broader legislative context and in the references made to the previous parliamentary debate in the Explanatory Memorandum.[26] This changed more significantly with the amendments adopted

[21] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019).
[22] European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI), OJ C 404, 6.10.2021, p. 129.
[23] Ibidem, para. 18: "Notes, with regard to the use of non-personal data by AI technologies, that the lawful use of copyrighted works and other subject matter and associated data, including pre-existing content, high-quality datasets and metadata, needs to be assessed in the light of the existing rules on limitations and exceptions to copyright protection[...]"
[24] Ibidem, Recital D: "whereas AI technologies may render the traceability of IPRs and their application to AI-generated output difficult, thus preventing human creators whose original work is used to power such technologies from being fairly remunerated."
[25] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final.
[26] Ibidem, §1.1.

by the European Parliament on 14 June 2023,[27] through which the original proposal was extensively revised and supplemented in light of the rapid evolution of AI technologies and, in particular, GPAI.

This regulatory process culminated in June 2024 with the adoption of Regulation (EU) 2024/1689, commonly known as the AI Act. The legislative text addresses copyright only indirectly, with a single provision, Article 53, and its related Recitals (104–108) specifically engaging with the issue. Nonetheless, their inclusion, together with the accompanying Codes of Practice, is highly relevant for the interpretation of the existing copyright framework and for understanding the underlying rationale of the EU legislator. These aspects will be examined in greater detail in the following sections.

The last evolution on the topic is the European Parliament resolution of 10 March 2026 on copyright and generative artificial intelligence.[28] It considers the ambiguities of the current legislative landscape of copyright in relation to G-AI training and calls the Commission to perform further investigation among stakeholders in order to propose a more suitable and clear framework.

2. G-AI and Its Implications for Copyright

At this stage, it becomes necessary to assess how the legislative frameworks outlined above interact with G-AI training in the field of copyright law. The starting point is to clarify in what ways the training of G-AI systems may affect copyright-protected works and the rights attached to them.

As noted above, G-AI training requires the large-scale ingestion of content in order to generate new outputs, and such training datasets may also include copyright-protected materials. It is therefore essential to identify which rights may be

[27] European Parliament, Amendments adopted on 14 June 2023 on the proposal for a regulation laying down harmonised rules on artificial intelligence, P9_TA(2023)0236.
[28] European Parliament, Committee on Legal Affairs, Report on copyright and generative artificial intelligence — opportunities and challenges, P10_TA(2026)0066.

implicated in this process under the European legal framework, and whether any exceptions or limitations may apply to these uses.

2.1. Rights Potentially Implicated by G-AI Training

G-AI may raise copyright issues in relation to both moral rights and economic rights. These concerns may arise at two distinct stages: the input phase, namely, the actual training of the model, and the output phase, which concerns the interaction between the model and the end user. Although both phases may be problematic from a copyright perspective, this thesis is limited to the potential infringements arising during the input phase. Furthermore, the current section will solely assess the impact on exclusive rights, since these are the rights more specifically implicated in the input phase.[29]

[29] While it is evident that moral rights may also be implicated during the output phase, it is noteworthy to acknowledge that certain legal scholars contend that issues in this field may also arise during the input phase. Firstly, it should be noted that moral rights have not been harmonised at EU level. As stated in Recital 19 of the Infosoc Directive, "The moral rights of rightholders should be exercised according to the legislation of the Member States and the provisions of the Berne Convention for the Protection of Literary and Artistic Works, of the WIPO Copyright Treaty and of the WIPO Performances and Phonograms Treaty. Such moral rights remain outside the scope of this Directive." Nevertheless, moral rights are recognised in international copyright law and in the domestic laws of Member States. They generally include the right of disclosure, the right of attribution, and the right of integrity; in some legal systems, they also include a right of withdrawal or retraction.
In particular, the right of integrity may be especially relevant in this context. This right protects the work against distortion, mutilation, modification, or other derogatory treatment capable of prejudicing the honour or reputation of the author. This issue may be particularly significant in Member States such as France and Belgium, where the right of integrity is understood according to a more subjective approach. In such legal systems, it is largely the author who determines whether a certain alteration amounts to derogatory treatment. By contrast, under a more objective approach, the existence of prejudice must be established on objective grounds.
The relevance of moral rights in the context of AI training stems from the fact that such training does not merely involve the digitisation of a work, which in itself does not normally involve the right of integrity, but rather its computational processing for a new technological purpose. In this sense, even in the absence of any communication to the public at the output stage, it could be argued that the use of a work in the training process may, in certain circumstances, be perceived by the author as a derogatory treatment of the work. Against this background of fragmentation, part of the scholarship has emphasised the need for a response at EU level, for example through the adoption of a specific exception or limitation.
It is also already worth highlighting here, as noted in the European Commission's *Study on Copyright and New Technologies: Copyright, Data Management and Artificial Intelligence*, the fact that the TDM exception (see §2.2.2.) is conceived only in relation to economic rights does not exclude the possibility that authors may oppose the use of their works as AI training inputs on the basis of moral rights.
For reference, see J. DREXL et al., "Artificial Intelligence and Intellectual Property Law-Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate", Max Planck Institute for Innovation Competition Research Paper 2021 No. 21-10, p. 12; and, European Commission: Directorate-General for Communications Networks, Content and Technology, Technopolis Group, UCLouvain, IMC University of Applied Sciences Krems, Crowell&Moring and

With regard to the economic exploitation of works and other protected subject matter, EU law provides a significant degree of harmonisation through the Infosoc Directive. However, such harmonisation covers only certain specific rights, namely the reproduction right (Article 2), the right of communication to the public (Article 3), and the distribution right (Article 4). Alongside these harmonised rights, EU law also recognises more sector-specific forms of protection that may prove relevant in the context of G-AI training. Given their importance for the present analysis, the following sections will focus especially on the reproduction right (in relation to Article 3 InfoSoc, databases, computer programmes and press publications) and on extraction right for *sui generis* databases under Directive 96/9/EC.[30]

2.1.1. Reproduction Right

General provision: Article 2 Infosoc

Among these rights, the reproduction right appears to be the one most directly engaged by G-AI training. According to the case law of the Court of Justice, the protection conferred by Article 2 of the Infosoc Directive must be interpreted broadly, so as to encompass direct or indirect, temporary or permanent

Philippe Rixhon Associates, p. 189. Study on copyright and new technologies – Copyright data management and artificial intelligence, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2759/570559

[30] Although proposed in relation to text and data mining rather than to G-AI training, the reconstruction that follows is pertinent here, since training is articulated in processing steps that are at least analogous to those of TDM (infra, §2.2.2.1). Servanzi analyses data mining as a sequence of logical steps — the creation of a working copy of the source, the organisation of the data, the search for correlations among them, and the use of the results — and reads the statutory definition (Article 70-ter(2) LDA, transposing Article 2(2) CDSM) as identifying the analytical activity itself, namely the organisation of the data and the search for correlations, as the definitional core of text and data mining, distinct both from the ancillary reproductions that precede it and from the subsequent use of its results. On this reading, the reservation of rights (provided by Article 70-quater LDA, transposing Article 4 CDSM Directive, *infra* Chapter II §2.2.) would confer on the rightholder a power to control the analytical activity as such, which the author classifies, depending on the case, either as a faculty of exploitation of the work or as an autonomous related right. This thesis does not adopt that reconstruction: consistent with the prevailing reading, the analytical activity is not treated here as an independently restricted act, since copyright and the sui generis right are engaged by the acts of reproduction and extraction, while the analysis as such remains free; the TDM exception correspondingly operates on the reproduction and extraction rights. Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §§1–2, pp. 1150–1153 and 1156–1157.

reproductions, by any means and in any form, in whole or in part.[31] In order to assess whether G-AI training might involve acts of reproduction under Article 2 it is first necessary to consider, in general terms, how the training process operates and whether it falls within the notion of an act of reproduction.

During the training process, the model ingests vast quantities of data in order to analyse them and learn from them. In practical terms, this means that the relevant content is collected, copied, stored, and pre-processed so that it can be analysed computationally by the system as numerical parameters (weights) reflecting the relationships between the units of the ingested data, whether textual, visual, or otherwise. The result of this statistical process is the model. There is no doubt that the acts of copying and storing at the beginning of the process constitute a form of reproduction within the meaning of Article 2 of the InfoSoc Directive. This conclusion is further supported at the international level by the Agreed Statements concerning the WIPO Copyright Treaty, approved on behalf of the European Community by Council Decision 2000/278/EC, which clarify that the reproduction right fully applies in the digital environment and that the storage of a protected work in digital form in an electronic medium constitutes a reproduction within the meaning of Article 9 of the Berne Convention.[32]

However, interpretative problems arise in relation to the nature of G-AI models. Both scholarship and jurisprudence are divided, arguing, on the one side, that the model memorizes the training materials and it is to be considered a form of reproduction, even if indirect, on the other side, that it is merely a statistical process and it does not involve acts of reproduction. At first sight, the controversy about

[31] CJEU, Infopaq International A/S v Danske Dagblades Forening, C-5/08, EU:C:2009:465, paras. 42 et seq. For literature, M. Bertani, Diritto d'autore europeo, Giappichelli, Torino, 2011 (Quaderni di AIDA, 21), cap. IV, §3; v. anche M. Bertani, A. Cogo, P. Fabbio, A. Genovese, A. Ottolia, Lineamenti di diritto industriale. Concorrenza e proprietà intellettuale, CEDAM, Padova, 2024, Parte VII, cap. IV, §3.1.

[32] Agreed Statements concerning the WIPO Copyright Treaty, adopted by the Diplomatic Conference on 20 December 1996. The statement clarifies that the reproduction right under Article 9 of the Berne Convention fully applies in the digital environment, and that storage of a protected work in digital form in an electronic medium constitutes reproduction within the meaning of that provision; see also Article 1(4) WCT: "Contracting Parties shall comply with Articles 1 to 21 and the Appendix of the Berne Convention".

this matter may appear to concern a predominantly technical question, namely whether a model "stores" the data on which it has been trained. Upon closer inspection, however, the disagreement is also conceptual and legal, since it depends on what is taken to be legally relevant for the purposes of copyright analysis. Recent scholarship shows that the divergence does not merely concern the empirical functioning of models, but also the appropriate focus of the inquiry.

A first line of argument, associated in particular with Cooper and Grimmelmann,[33] places emphasis on the model's capacity to regenerate protected expression. In their account, memorisation should not be reduced to the simple fact that a user is able to extract verbatim material through adversarial prompting. Rather, memorisation is said to occur when it is possible to reconstruct from the model a near-exact copy of a substantial portion of a training work. On this basis, they distinguish memorisation from related but different notions such as extraction, regurgitation, and reconstruction. Cooper and Grimmelmann do not present extraction, regurgitation, and reconstruction as different types of memorization; rather, they treat memorization as the underlying phenomenon, while regurgitation is a symptom/manifestation of it; extraction is the means of revealing it; reconstruction is the criterion that shows memorization is present.[34]

Their central claim is that memorisation *is a property of the model itself*, arising during training, whereas regurgitation is only one possible manifestation of that underlying condition at generation stage. In this sense, the model's ability to *reconstruct* exact or near-exact outputs is treated not merely as an accidental downstream event, but as evidence that protected expression is in some legally

[33] Cooper, A. F., & Grimmelmann, J. (2024). The files are in the computer: Copyright, memorization, and generative AI. arXiv preprint arXiv:2404.12590. https://arxiv.org/abs/2404.12590

[34] Ibidem, "It is helpful to distinguish three related senses in which a model might colloquially be said to have "memorized" its training data. [...]Most narrowly, when a user intentionally and successfully prompts a model to generate an output that is an exact or near-exact copy of a piece of training data, 62 that is extraction; More broadly, when a model generates an output that is an exact or near-exact copy of piece of training data (whether or not the user intentionally prompted the model with that goal), that is regurgitation; Most broadly of all, when an exact or near-exact copy of a piece of training data can be reconstructed by examining the model "through any means," that is memorization We will use the term reconstruction to refer to these different means, which can include but are not limited to prompting."

meaningful way present within the model. Cooper and Grimmelmann therefore argue that, where memorisation in this sense occurs, the model may itself qualify as a "copy" of the memorised training material for copyright purposes.

This first approach has two important implications. First, it shifts the analysis away from a narrow focus on literal file storage and towards a broader understanding of reproduction capable of encompassing model parameters and internal representations. Second, it weakens the argument that memorisation should matter only when harmful outputs are actually generated. If memorised aspects of a work are already present in the model, then the copyright issue arises at the level of training and model construction, not only at the stage of public-facing outputs. In this perspective, the model's occasional ability to reproduce protected content is not simply a symptom of overfitting or an unusual malfunction, but a sign that the work has been retained in a form that may be legally relevant[35].

This position also finds support in German jurisprudence. In particular, the judgment of the Munich Regional Court in GEMA v OpenAI constitutes a significant development, even though it is only a first-instance decision and is currently under appeal. The case concerns several issues that will also be relevant in other parts of this thesis. For the purposes of the present section, however, the judgment is especially important because of its treasoning on memorisation.

Briefly as to the facts, GEMA, the German collective management organisation for music rights, brought proceedings against OpenAI alleging that the defendant's language model contained copyright-protected song lyrics. GEMA also argued that the relevant works had been subject to an opt-out, although that issue is not central to the present section and will not be considered by the Court, since it is absorbed by the other and previous arguments.

[35] For further scholarship Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, p. 45.
https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)774095

As regards memorisation, the Munich court held that the decisive point for the purposes of reproduction was that the song lyrics used as training data were reproducibly incorporated into the model and thus embodied within it.[36] The Court's reasoning is based on a broad understanding of reproduction, drawing on CJEU case law[37] and in particular on the idea that reproduction does not require direct perceptibility, since indirect perceptibility may suffice. In the Court's view, this is precisely what occurs in AI training: even if outputs are generated probabilistically, the possibility of reproducing the protected work remains sufficient for a finding of reproduction. The Court further clarified that limited perceptibility does not exclude reproduction, but may at most be relevant to the separate question whether the work is publicly accessible, which is not a condition of the reproduction right.[38]

In this respect, the judgment appears broadly consistent with the approach advanced by Cooper and Grimmelmann, insofar as it places particular emphasis on the model's ability to reproduce or regenerate protected expression as evidence of legally relevant retention within the model. The decision is also noteworthy for its position on the applicability of the TDM exception to AI training, an issue that will be addressed in the following sections.

A different line of argument, developed by Leistner and Antoine, contests the conclusion that trained AI models contain reproductions of copyright-protected training material from a legal-systematic perspective. Their starting point is that, under EU copyright law, an exclusive right can only be infringed if the relevant use act concerns an element that is itself capable of copyright protection. Accordingly, the harmonised concept of a 'work' — as defined by the CJEU —

[36] GEMA v OpenAI, Landgericht München I, Az. 42 O 14139/24, 11 November 2025, para. 186 (original German: "dass die Liedtexte, die als Trainingsdaten dienten, im Modell reproduzierbar enthalten und somit verkörpert sind"; English translation: "the song lyrics used as training data are reproducibly incorporated into the model and are thus embodied within it").
[37] CJEU, Copydan Båndkopi v Nokia Danmark A/S, C-463/12, EU:C:2015:144, para. 35 (decided in the context of private copying, but addressing the concept of reproduction).
[38] GEMA v OpenAI, Landgericht München I, Az. 42 O 14139/24, 11 November 2025, paras. 188–189.

sets a decisive threshold: beyond the requirement of the author's own intellectual creation, the existence of a subject matter identifiable with sufficient precision and objectivity is required.[39] This basic prerequisite is not fulfilled with regard to a potential reproduction within a trained AI model. Even for an expert, it is not possible to detect and identify a particular work within the model's parameters. Works used for training can only 'reappear' as the result of a particular user prompt — an outcome that is merely the expression of a probabilistic effect which cannot be predicted or traced back with sufficient certainty. From a legal perspective, the causal link that exists at the technical level is thus interrupted by the user's action, whose results cannot be foreseen with sufficient certainty. This situation, the authors argue, cannot be compared to a reproduction that requires certain technical means in order to be perceivable, since the latter is at least precisely predetermined as to its conditions and effects.[40] In addition to this reasoning, Leistner and Antoine also provide a more systemic interpretation relation to the TDM exception of Article 3 and 4 CDSM Directive which is assessed in §2.2.2.2.

What appears to differ most sharply between these two approaches is the level at which the copyright analysis is conducted. Cooper and Grimmelmann, and the Munich court in GEMA, assign decisive weight to the model's demonstrable capacity to regenerate protected expression, treating that capacity as evidence of a legally relevant retention within the model's parameters. Leistner and Antoine, by contrast, argue that the analysis must begin one step earlier: not with what the model can produce, but with whether a copyright-protected work is identifiable with sufficient precision and objectivity within the model itself. Since this condition — which the CJEU requires at the infringement level — cannot be satisfied in relation to AI model parameters, the conclusion that the model constitutes a reproduction does not follow, even where occasional near-exact

[39] CJEU, Cofemel — Sociedade de Vestuário SA v G-Star Raw CV, C-683/17, EU:C:2019:721, para. 29; CJEU, Brompton Bicycle Ltd v Chedech/Get2Get, C-833/18, EU:C:2020:461, para. 22

[40] Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §III.2. https://doi.org/10.1093/grurint/ikaf114

outputs can be generated. The disagreement is therefore not merely about technical facts, but about which legal question should be asked first: whether the model can reproduce, or whether the model contains something that copyright law can recognise as a work.

Support for the position that trained AI models do not contain reproductions of their training material can also be found in Getty Images (US) Inc and Others v Stability AI Ltd, decided by the English High Court.[41] That case concerned, within the UK copyright framework, the notion of an "infringing copy" or "infringing article".[42] Getty Images, a company operating in the stock-image sector, alleged that Stability AI had used millions of its images without consent in order to train Stable Diffusion, the defendant's AI model. What is particularly relevant for the present section is the Court's characterisation of the model itself. The judge, Joanna Smith, addressed the question whether an AI model resulting from a training process involving exposure of model weights to infringing copies could itself qualify as an infringing copy.[43] The answer was negative. The Court stated that the model weights were not themselves an infringing copy and did not store an infringing copy, but were rather "the product of the patterns and features which they have learnt over time during the training process." It further stated that the fact that the model's development involved acts of reproduction of copyright works was not, in itself, sufficient to make the model an infringing copy.[44]

[41] Although this decision was rendered under UK copyright law, it is not without relevance to the EU law analysis. As the Court of Appeal clarified in TuneIn Inc v Warner Music UK Ltd, CJEU judgments rendered before 31 December 2020 constitute retained EU case law and continue to bind UK courts, while post-Brexit CJEU judgments may be taken into account by UK courts when relevant.[2] Accordingly, the UK courts' approach to questions of EU copyright law is not entirely disconnected from the EU framework, and the Getty Images decision — while not binding as a matter of EU law — nonetheless illustrates that the conclusion that AI model parameters constitute reproductions of training material is not self-evident from a judicial perspective. For reference: TuneIn Inc v Warner Music UK Ltd [2021] EWCA Civ 441, paras. 73–75 and 89 (Arnold LJ).

[42] Copyright, Designs and Patents Act 1988 (UK), sections 22 and 23 read with section 27(3).

[43] Getty Images (US) Inc and Others v Stability AI Ltd [2025] EWHC 38 (Ch), para. 599: "Taking the specific facts with which I am concerned, is an AI model which derives or results from a training process involving the exposure of model weights to infringing copies itself an infringing copy?"

[44] Ibidem, para. 600.

However, at present, it still appears difficult to regard either of these approaches as conclusively prevailing. The divergence between them does not seem to derive solely from incomplete technical knowledge, but also from a deeper methodological choice concerning the proper object of legal evaluation: whether priority should be given to model behaviour and regenerative capacity, or instead to internal architecture and the distinction between storage and learning.

At the same time, recent case law suggests that, at least within the EU legal environment, an emerging tendency may be identified in favour of treating memorisation in generative AI models as potentially relevant to the reproduction right.

A further consideration, which bears directly on the legal assessment above, concerns recent technical developments aimed at reducing or eliminating memorisation altogether. Indeed, in the context of the "reproducibility" criteria, it is imperative to emphasise that contemporary developments in G-AI models are directed towards the prevention or, at the very least, the minimisation of the direct replication of training materials during the output phase. It is possible to achieve this objective through the curation of datasets, that is to say, the preparation of the dataset for training prior to the commencement of the training process. Alternatively, this objective can be pursued through the establishment of training objectives, that is to say, the focus on the architecture of the model during the training process. Finally, this objective can be achieved through the implementation of post-training procedures.[45] The methods contained within these categories are numerous, and the following examples will illustrate each category. With regard to the first, the technique is concerned with the process of deduplicating the data to be ingested. This method is intended to ensure that, in terms of statistical probability, the "formulation" of the sentence is rendered less

[45]Sakarvadia, M., Ajith, A., Khan, A., Hudson, N., Geniesse, C., Chard, K., Yang, Y., Foster, I., & Mahoney, M. W. (2024). Mitigating memorization in language models. arXiv preprint arXiv:2410.02159.

likely.[46] The second approach, on the other hand, focuses on tokenization and pre-training phases. The initial term pertains to the conversion of text/content into a word unit, single words, symbols, and other linguistic elements (referred to as "tokens") that can be understood by the machine. The subsequent term involves the reading, analysis, and generation of statistical patterns from the tokenised data. In a standard training process, all tokens would be used, and statistical predictions would be made on them all. However, the proposed architecture, which aims to minimise memorisation, excludes randomly selected subsets of tokens from the training.[47] In both cases, the result is a model that is incapable of reproducing the ingested content, or which only has a very limited capacity to do so.

The third method is particularly relevant for models that have already been trained, and it generally corresponds to the process of machine unlearning. The objective of this technique is to eliminate the impact of training data on a trained model. The distinguishing factor is that the model is already in existence and undergoes a process that facilitates memorisation without compromising, if the process is well managed, the model's performance. The result of this process is a new model that exhibits either reduced or no memorisation.[48]

The relevance of these technical developments is further acknowledged at the regulatory level. Measure 1.4 of the Code of Practice for GPAI requires signatories to implement appropriate and proportionate technical safeguards to prevent their models from generating outputs that reproduce training content protected by EU copyright law in an infringing manner.[49] Although Measure 1.4 operates at the output stage — addressing the risk of infringing generation rather than the training

[46] Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2022). Deduplicating training data makes language models better. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022).
[47] Hans, A., Wen, Y., Jain, N., Kirchenbauer, J., Kazemi, H., Singhania, P., Singh, S., Somepalli, G., Geiping, J., Bhatele, A., & Goldstein, T. (2024). Be like a goldfish, don't memorize! Mitigating memorization in generative LLMs. In Proceedings of NeurIPS 2024.
[48] Sakarvadia, M., Ajith, A., Khan, A., Hudson, N., Geniesse, C., Chard, K., Yang, Y., Foster, I., & Mahoney, M. W. (2024). Mitigating memorization in language models. arXiv preprint arXiv:2410.02159.
[49] Code of Practice for General-Purpose AI Models, Copyright Chapter (2025), Measure 1.4(1)(a).

process itself — it implicitly recognises that the capacity of a model to reproduce training content is a legally relevant concern that providers must actively mitigate. In this sense, the technical measures described above — in particular training architecture modifications and machine unlearning — may also serve as instruments of compliance with the obligations imposed by the Code of Practice, insofar as they reduce or eliminate the model's capacity to regenerate protected content.

This has significant implications for the reproduction right analysis: if memorisation can be technically prevented, the legal concern identified in GEMA would lose its factual basis.[50]

In conclusion, there are no doubts that the training of G-AI models involves acts of reproduction in relation to the intermediate copies made for the content ingestion, but whether G-AI models constitute a form of reproduction itself, at least for the time being, remains an open and evolving matter.

Copyright-protected databases

The relevance of the reproduction right also extends to the copyright protection of databases. Under Article 3 of Directive 96/9/EC, databases are protected by copyright where, by reason of the selection or arrangement of their contents, they constitute the author's own intellectual creation. Article 5(a) of the Directive further provides that the author enjoys, among other exclusive rights, the right to reproduce the database, by any means and in any form, in whole or in part. It should be emphasised, however, that this protection concerns the database as such, and not the individual contents contained within it, regardless of whether those contents are themselves protected.[51]

[50] For legal literature on memorization mitigation see EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 155.
[51] Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press. https://doi.org/10.1093/oso/9780198858591.001.0001

Computer programmes

Moreover, Directive 2009/24/EC (Software Directive) provides protection for computer programs that is equivalent to that for literary works.[52] The object of the protection is the expression of the software in any form (the source code as well as the object code).[53] Article 4 of the Software Directive provides protection for computer programmes from both reproduction and alteration. Indeed, G-AI also involves the ingestion of training materials to learn the "programming" languages, which makes the reproduction in this field also significant. Alteration, instead, is irrelevant for the input phase of G-AI.[54]

Press publications

Finally, it should be mentioned that press publishers also enjoy a neighbouring right in relation to press publications, according to the definition set out in CDSM Directive as "[...] a collection composed mainly of literary works of a journalistic nature [...]".[55] This right, without affecting the author's rights on the work, grants the publisher the exclusive right of reproduction and of making the publication available to the public.[56] The former, the reproduction right, is of particular pertinence for this dissertation, given the substantial role played by press publications in the domain of G-AI training. Indeed, it is precisely this content that ensures the continual updating of the model.

---

[52] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, OJ L 111, 5.5.2009, pp. 16–22, Article 1(1). "[...] Member States shall protect computer programs, by copyright, as literary works within the meaning of the Berne Convention for the Protection of Literary and Artistic Works"

[53] CJEU, Bezpečnostní softwarová asociace — Svaz softwarové ochrany v Ministerstvo kultury, C-393/09, EU:C:2010:816.

[54] Alteration, within the meaning of Article 4(1)(b) of Directive 2009/24/EC, encompasses any adaptation, translation, arrangement or other modification of a computer program, as well as the reproduction of the results thereof. Unlike the reproduction right, which is engaged by the mere copying of the program, alteration requires a transformation of the program's expressive content. In the context of G-AI training, alteration is not directly engaged at the input stage, since the training process does not modify the software ingested as training material but rather analyses it computationally.

[55] 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, OJ L 130, 17.5.2019, pp. 92–125, Article 2(4).

[56] Ibidem, Article 15.

2.1.2. Extraction Right

Directive 96/9/EC provides a further layer of protection for databases in the form of a *sui generis* right. Unlike copyright protection under Article 3, this right applies where the database is the result of a substantial qualitative or quantitative investment, regardless of whether its selection or arrangement meets the originality threshold required for copyright protection. In any case, a database could be protected both by copyright and by *sui generis* right.

The key provision is Article 7, which grants the maker of the database the right to prevent extraction and re-utilisation of all or a substantial part of its contents. Article 7(2) defines extraction as "the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form". The rationale of this protection lies not primarily in safeguarding the contents as such, but in protecting the investment made in obtaining, verifying, or presenting them.[57] For this reason, the notion of extraction has been interpreted broadly, as to cover acts of appropriation that may prejudice that investment. In particular, the transfer and storage of all or a substantial part of the contents of a protected database in another medium may fall within the scope of Article 7.[58]

This aspect is particularly relevant where training datasets are compiled through the transfer of substantial parts of protected databases, since in such cases the legal issue may concern not only copyright in the database structure, but also the *sui generis* right protecting the underlying investment.

[57] Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press. https://doi.org/10.1093/oso/9780198858591.001.0001; M. Bertani, 'Banche dati ed appropriazione delle informazioni', in Europa e diritto privato, 2006, §2; v. anche M. Bertani, A. Cogo, P. Fabbio, A. Genovese, A. Ottolia, Lineamenti di diritto industriale. Concorrenza e proprietà intellettuale, CEDAM, Padova, 2024, Parte VII, cap. II, §18.

[58] CJEU, Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg, C-304/07, EU:C:2008:552.

Against this background, G-AI training (as described above) may also entail acts of extraction, insofar as it involves the transfer and incorporation of database contents into a different technical environment for the purposes of training.

2.2. Assessing the Applicability of Exceptions

Both the Infosoc Directive and the CDSM Directive establish a system of exceptions and limitations to the exclusive rights conferred under EU copyright law. On the one hand, Article 5(1) of the Infosoc Directive provides for an exception covering certain temporary acts of reproduction (only reproductions, not extractions). On the other hand, the CDSM Directive introduces specific exceptions for text and data mining, distinguishing between uses for scientific research (Article 3) and uses for other purposes (Article 4).

2.2.1. Temporary Acts of Reproduction

It is noteworthy that the CDSM directive also names the temporary exception in relation to TDM. In particular, Recital 9 refers to Article 5(1) of the Infosoc Directive stating that the latter keeps applying to TDM techniques that do not involve making copies beyond the scope of exception.[59] However, the Infosoc Directive, in Recital 18 specifies that the fulfilment of Article 5(1) requirements by TDM techniques is not certain.[60] The following analysis assesses this uncertainty. As well, it is of particular relevance to point out that in relation to traditional TDM literature is consistent in considering the temporary acts exception inapplicable.[61]

In order for Article 5(1) to apply, all of its conditions must be cumulatively fulfilled. More specifically, there are five conditions: the act must be temporary, namely the process must be automated and must not require human intervention

[59]EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 49.
[60] Ibidem.
[61] L. Mansani, 'Le eccezioni per estrazione di testo e dati, didattica e conservazione del patrimonio culturale', in AIDA, 2019, §2, observing that the technique entails permanent reproductions.

for the deleting; it must be transient or incidental, namely the reproduction must not last more than necessary for the technological process, or it must be only ancillary to the main process and not have an independent purpose; it must constitute an integral and essential part of a technological process, meaning that the process must be incapable of working without that reproduction; the sole purpose of the act of reproduction must be either a transmission in a network between third parties by an intermediary or a lawful use of a work or other subject matter, namely, for the latter, there must be the rightholder authorization or the act must be not restricted by law;[62] finally, the act must not have independent economic significance, namely the reproduction must not have additional and separate economic profit.[63] The rationale underlying this exception is to exempt certain acts of temporary reproduction that are technically necessary for the functioning of digital networks and that lack autonomous economic value, such as browsing and caching.[64] In light of its derogatory character, the Court of Justice has consistently required these conditions to be interpreted strictly.[65]

As mentioned, the conditions are cumulative and the absence of one determines the inapplicability of the exception. With regard to G-AI training, the case of LAION v Robert Kneschke, which was decided by the District Court of Hamburg, provides an example that addresses this issue. In the case at hand, the Court decided that the downloading carried out by LAION, a non-profit organisation specialising in the development of open-source AI models and datasets (in this particular instance, the object was a dataset), was not merely ancillary to the technological process of training.[66] The Court stated that the acts of reproduction performed did

[62] Infosoc Directive, Recital 33.
[63] CJEU, Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others, C-360/13, EU:C:2014:1195, para. 22.
[64] M. Bertani, Diritto d'autore europeo, Giappichelli, Torino, 2011 (Quaderni di AIDA, 21), cap. IV, §8, lett. A.; EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 50; Sganga, C. (2021). The right of reproduction. In The Routledge handbook of EU copyright law (pp. 123–150). Routledge;
[65] CJEU, Infopaq International A/S v Danske Dagblades Forening, C-5/08, EU:C:2009:465; CJEU, Football Association Premier League Ltd and Others v QC Leisure and Others, C-403/08, EU:C:2011:631.
[66] District Court of Hamburg (Landgericht Hamburg), judgment of 27 September 2024, 310 O 227/23, paras. 63–64.

not fulfil the criteria of being 'temporary' and 'transient or incidental'.[67] While it is important to refer to this judgment, as it is the only one addressing the problem, it must be noted that in the case at issue, the process performed by LAION was about creating datasets for AI training (reproduction, filtering and cleaning of photographs in a semi-structured way), and not about the training of the AI model.[68] It is indisputable that in order to construct its dataset, LAION also employed analysis techniques to associate the images with the related texts. Furthermore, it is equally true that the creation of the dataset constitutes one of the requisite stages of the G-AI training. These elements provide a sound justification for the extension of the TDM exception, as stipulated in Article 3 CDSM (*infra*), to LAION's activities as determined by the Court. Nonetheless, in the case at hand, the training of the model is only subsequent and incidental.

For this reason, the Court's reasoning on the inapplicability of Article 5(1) Infosoc Directive cannot automatically be transposed to the training of a G-AI model as such. Indeed, reproductions can hardly be regarded as temporary or transient where the very result of the process is a database in which those reproductions are stored. It should also be noted that it is not inherent to the training pipeline that intermediate reproductions be retained once training has been completed: whether such copies are preserved or deleted depends on the provider's storage practices.[69] If the trained model is not itself regarded as a reproduction — a question examined in the preceding section on reproduction (§2.1.1.) — the temporary character of at least some training-related reproductions cannot be excluded as a matter of technical design. In any case, the question whether the remaining cumulative conditions of Article 5(1) would be satisfied in such a scenario remains unaddressed by national and EU courts, and given the significance of such a

[67] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office.
[68] Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §I and §II.1. https://doi.org/10.1093/grurint/ikaf114
[69] Sugimura, P., & Hartl, F. (2018). Building a reproducible machine learning pipeline. arXiv preprint arXiv:1810.04570, §2.1.

finding, an authoritative judicial assessment would be necessary before any firm conclusions can be drawn.[70]

2.2.2. Is TDM Exception Applicable to G-AI Training?

The applicability of the TDM exception to G-AI training is a debated topic in legal scholarship, jurisprudence and even at the institutional level. This section provides a systematic reading of the legislation and addresses the main points of contention. Indeed, the case Like Company v. Google pending before the CJEU is indicative of the ambiguity of the current framework since the third preliminary question specifically asks whether the reproduction made with the purpose of training a Large Language Model (LLM)-based[71] chatbot falls within the TDM exception of Articles 3 and 4 of the CDSM Directive.[72] While the present paragraph provides a general introduction to the exception, the following sections (§2.2.2.1. and §2.2.2.2.) analyse whether and to what extent G-AI training falls within the EU definition of TDM and address some structural problems subsequent to the definitional inquiry (memorization and three-step test compliance). Moreover, the specific requirements and challenges of Articles 3 and 4 (subjects, lawful access, opt-out, transparency and enforcement) are assessed in Chapter II.

The CDSM Directive in Article 2(2) defines TDM as "[…] any automated analytical technique aimed at analysing text and data in digital form in order to

[70] The analysis of this section does not consider Retrieval Augmented Generation (RAG), which is a method primarily employed to maintain the currency of G-AI systems in relation to information that was not utilised during the training of the model, e.g. due to its recent nature. In practical terms, it integrates external materials, for example through an API that connects to the web, and elaborates them through the model, without actually incorporating them into it. This technique is increasingly utilised due to its capacity to minimise the frequency of model retraining. Whilst this method may be subject to differing evaluations in relation to Article 5(1) of the Infosoc Directive, it is not the focus of this dissertation as it does not constitute a form of G-AI training. For further analysis, see Dornis, T.W., Lucchi, N. Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of the Referral in Like Company v. Google. IIC 56, 1800–1840 (2025). https://doi.org/10.1007/s40319-025-01649-7. §2.1.

[71] LLMs are neural network models, mostly transformers, specifically designed for language acquisition and processing. They constitute a form of G-AI model, which is a broader term encompassing AI models also designed to generate images, audio, and video.

[72] CJEU, Case C-250/25, Like Company v Google Ireland Limited, request for a preliminary ruling, third question: "[...] does such reproduction of lawfully accessible works fall within the exception provided for in Article 4 of Directive 2019/790, which ensures free use for the purposes of text and data mining?

generate information which includes, but is not limited to, patterns, trends and correlations." This definition is very broad and does not substantially differ from the one originally proposed by the European Commission in the legislative proposal.[73]

The rationale behind the introduction of the new exception can already be traced in the Commission's 2015 Communication A Digital Single Market Strategy for Europe, where it was observed that "Innovation in research for both non-commercial and commercial purposes, based on the use of text and data mining (e.g. copying of text and datasets in search of significant correlations or occurrences) may be hampered because of an unclear legal framework and divergent approaches at national level. The need for greater legal certainty to enable researchers and educational institutions to make wider use of copyright-protected material, including across borders, so that they can benefit from the potential of these technologies and from cross-border collaboration will be assessed […]."[74] Indeed, TDM represents a fundamental practice in the context of machine learning (ML), a computational method based on data and algorithms, which is used to find patterns and trends in large datasets or to make predictions, essential in the field of data science. As noted by the Commission, this practice has become increasingly important across a wide range of scientific and technological fields.

However, the TDM exception underwent several modifications during the legislative process. While the Commission's proposal contained a single provision on TDM (Article 3), the final text of Directive (EU) 2019/790 introduced two distinct provisions: Article 3, dedicated to scientific research purposes, and Article 4, which provides for a broader exception covering any purpose. The two

[73] Article 2(2) of the Commission Proposal, COM/2021/206 final, defined TDM as "any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations."
[74] European Commission, Communication: A Digital Single Market Strategy for Europe, COM(2015) 192 final, §2.4.

provisions differ in several respects. First, while both the provisions admit acts of reproduction within Article 2 Infosoc Directive, databases, and press publications, only Article 4 encompasses reproductions of computer programs. Second, a subjective perspective: Article 3 applies only to research organisations and cultural heritage institutions, whereas Article 4 is not limited to specific beneficiaries and therefore applies to all users. Third, with regard to the purposes covered: Article 3 only permits text and data mining for scientific research purposes, while Article 4 allows text and data mining for any purpose. Finally, Article 4 is subject to a specific condition: it applies only where the use of works and other subject matter has not been expressly reserved by the rightholders through an opt-out mechanism.[75] Despite these differences, both provisions share some common requirements. In particular, they apply only where the content has been lawfully accessed, and in both cases the retention of reproductions and extractions is limited. However, in relation to the latter, the condition is slightly different, while Article 3 has no temporal limitation, but the retention must be justified by scientific research purposes (even verification of the results is a possible justification); Article 4 is much more restrictive: the retention cannot extend beyond what is necessary for the purposes of text and data mining (*infra*, §2.2.2.2.)[76]

The objections concerning the applicability of the exception relate to different elements, which can be categorised as follows: firstly, the ones that relate to the functioning and purpose of the TDM and G-AI training itself;[77] secondly, the ones relating to the issue of memorisation (*infra*, §2.2.2.2.) in light of the retention limit

[75] CDSM Directive, Article 4(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 rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

[76] Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press. https://doi.org/10.1093/oso/9780198858591.001.0001

[77] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs.

laid down in Articles 4;[78] and, finally, the ones relating to the compliance with the three-step test of Article 5(5) of the Infosoc Directive.[79] With regard to the latter issue, it is noteworthy that CDSM Directive itself, under Article 7, requires the exceptions to be in compliance with Article 5(5) of the Infosoc Directive.

The following section is concerned with the first and preliminary issue, namely whether the definition set out in Article 2(2) may be regarded as applicable, in whole or in part, to G-AI training. Then, § 2.2.2.2. addresses the subsequent structural questions, namely whether the retention limits under Articles 3 and 4 can be reconciled with memorisation and whether, and to what extent, the application of the TDM exception to G-AI training is compatible with the three-step test.

2.2.2.1. Is G-AI Training TDM?

The definition of TDM encompasses three elements: it is (a) an automated analytical technique, (b) built to analyse text and data and (c) with the purpose of generating information. The discourse pertaining to the applicability of this definition to G-AI training is centred on two key areas. Firstly, it draws attention to the technical differences between G-AI training and TDM. Secondly, it focuses on the rationale behind the provision.

Many scholars highlight that G-AI training is not designed simply to analyse and find information; rather, it is designed to synthesise new expressions.[80] According to this view, TDM is a data science method, whereas G-AI is not. More specifically, both methods ingest large-scale data, but they are built to generate

[78] Cooper, A. F., & Grimmelmann, J. (2024). The files are in the computer: Copyright, memorization, and generative AI. arXiv preprint arXiv:2404.12590. https://arxiv.org/abs/2404.12590; for jurisprudence, see GEMA v OpenAI, Landgericht München I, Az. 42 O 14139/24, 11 November 2025.

[79] CDSM Directive, Article 7, read in conjunction with Article 5(5) of the InfoSoc Directive: "The exceptions and limitations provided for in paragraphs 1, 2, 3 and 4 shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder." This constitutes a transposition of the three-step test of the Berne Convention; while the former operates as an interpretative criterion, the latter constitutes a treaty obligation.

[80] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs.

fundamentally different outputs: a correlation (which might be a pattern, trend, etc) on the one hand, and a new expression on the other. The implications of this reasoning are not merely technical. While data science relates to the semantics of ingested material, G-AI training internalises the form of expression of creative works.[81] Those who support this interpretation argue that this is a far more problematic activity for copyright than TDM, since it involves the core of its protection.[82]

In addition to the technical arguments that have been presented, the intentions of the EU legislator are also employed as evidence. In particular, Recital 8 of CDSM Directive[83] is referred to in order to demonstrate that the exception pertains exclusively to information, new knowledge and discovering new trends.[84] G-AI training has been shown to result in the reproduction of content that is analogous to that which was utilised during the training process. This training, following this position, does not yield any novel knowledge.

There are a number of factors to be taken into consideration when assessing these critical points. Firstly, it is important to note that the G-AI training process is not

[81] A comparable distinction is drawn in Italian scholarship by Servanzi in relation to TDM techniques, but which is also relevant for G-AI training. The author differentiates between non-semantic data mining — where the machine extracts data forming part of the expressive form of a work without identifying the ideas or information conveyed through it — and semantic data mining, where the machine identifies and extracts the meaning conveyed by that form. The distinction is in turn related to two competing conceptions of the object of copyright protection: a formal theory, under which copyright protects the expressive form in itself, and a functional theory, under which protection attaches not to the form as such but to the combination of form and content, so that the use of pure form, severed from the content it conveys, does not amount to an exploitation of the work. The distinction offers a doctrinal vocabulary for the question, central to this thesis, of whether G-AI training engages protected expression or merely the unprotected informational substrate. Servanzi draws from it further consequences as to whether the reservation under Article 70-quater constitutes a formality prohibited by Article 5(2) of the Berne Convention; those consequences, however, presuppose that the object of the reserved power is the analytical activity — and, as regards reproductions, the functional theory's premise that the reproduction of pure form is not a reproduction of the work — a characterisation not adopted here (supra, §2.1), and they are accordingly not pursued. Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §§5–6, pp. 1166–1171.

[82] Ibidem.

[83] CDSM Directive, Recital 8: "New technologies enable the automated computational analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining. Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible"

[84] Dornis, T. W., & Lucchi, N. (2025). Generative AI and the scope of EU copyright law: A doctrinal analysis in light of the referral in Like Company v. Google. IIC — International Review of Intellectual Property and Competition Law, 56, 1800–1840. https://doi.org/10.1007/s40319-025-01649-7

a single, unified process; rather, it comprises numerous phases, from the development of the model to the actual generation of content.[85] Secondly, the AI Act's reference to the TDM exception for GPAI training suggests that the legislator intends to apply the exception to it as well.[86]

With regard to the first point, it is vital to understand the distinction between the G-AI model and the G-AI system. The former is the result of the training, and the latter is the interface that allows interaction between the model and the user. In other words, the model is the system's primary component (the engine).[87] The system is also capable of some activities. For example, it can connect to the web in order to provide a service that is always up-to-date without necessitating constant updates to the model, which is much more complicated and time-consuming.[88]

Some scholars argue that while the TDM exception may be applicable to the G-AI model training, its placement within an AI system — which often corresponds to its availability on the market — is not covered.[89] In support of this reading, they emphasise the use of the term "generate" in the definition of Article 2(2), which would allow for the extension of its meaning to G-AI training even if this was not the intention of the legislator at the time of adoption. The commercial deployment of the system would nonetheless fall outside the exception, also in light of the three-step test, which they content might not be respected.

The judges of the District Court of Hamburg upheld a slightly different position in the case LAION v. Robert Kneschke. It is acknowledged that there are different phases relating to G-AI (i.e. dataset creation, model training and subsequent use).

---

[85] European Copyright Society. (2025). Copyright and generative AI: Opinion of the European Copyright Society.
[86] AI Act, Article 53(1)(c) and (d), Recitals 104–108, and the Code of Practice for General-Purpose AI Models, Copyright Chapter (2025).
[87] Quintais, J. P. (2025). Generative AI, copyright and the AI Act. Computer Law & Security Review, 56, 106107. https://doi.org/10.1016/j.clsr.2025.106107
[88] See note 66 above (RAG).
[89] European Copyright Society. (2025). Copyright and generative AI: Opinion of the European Copyright Society.

However, it is stated that making it available to the public does not pose any problems for the application of TDM. In that particular case, it was the German transposition of Article 3 (Section 60b UrhG) to be applicable, and after denying the loss of the research purpose because of the potential use by other companies, it also implicitly admitted the applicability of the exception even following the publication.[90] The same Court also stated that AI training is not different from other TDM, and rejected the view that distinguishes between information 'hidden' in data and creative expression.[91]

It is evident that the obligation imposed on GPAI providers by the AI Act serves as an additional corroboration of a more favorable position. Article 53 imposes transparency obligations in both clause c) and clause d) of the first paragraph. The former requires the GPAI provider to *put in place a policy* in order to comply with EU copyright law. The latter stipulates the requirement for the disclosure of the content used for the training. In order to provide a more detailed overview, the initial obligation can be categorised into three distinct components: the formulation of the policy document, the maintenance of it up-to-date, and its subsequent release (*infra*, Chapter II, §3.)[92] It is very significant that it refers directly to the opt-out mechanism of Article 4(3). This, implicitly admits that TDM is performed during GPAI training, at least in part of it. Furthermore, even a more direct reference can be found in the correlated Recitals. Among others, Recital 105 is very clear, it states that "The development and training of such models require access to vast amounts of text, images, videos, and other data. Text and data mining techniques may be used extensively in this context for the retrieval and analysis of such content, which may be protected by copyright and related rights."[93] Such reference

[90] District Court of Hamburg (Landgericht Hamburg), judgment of 27 September 2024, 310 O 227/23, para. 107. It should be recalled that the LAION case concerned the creation of a dataset for AI training, not the training of a G-AI model as such.

[91] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office.

[92] European Copyright Society. (2025). Copyright and generative AI: Opinion of the European Copyright Society.

[93] AI Act, Recital 105, second sentence (it further clarifies that the use of the works must be respectful of the opt-out expressed by the rightholders.)

appears as a legislative confirmation of the applicability of the exception. A counterargument is based on Recital 108, which states that the "Regulation does not affect the enforcement of copyright rules as provided for under Union law." Some scholars argue that even if the AI Act refers to the exception it is not an implicit confirmation of its applicability. The TDM remains as it was conceived in the CDSM (eventually, in their opinion, as not entailing G-AI training) and, if it is used in a training proceeding it has to be acknowledged through the transparency document. Namely, Article 53 is procedural in nature and it "[...]does not confer new rights or extend the substantive scope of copyright exceptions under EU law."[94]

In conclusion, although the issue remains debated in legal scholarship, G-AI training appears to incorporate, at least in part, techniques that may be qualified as text and data mining.

2.2.2.2. Structural Limits of the TDM Exception in Relation to G-AI Training

Memorisation and compliance with the three-step test give rise to structural questions concerning the application of the exception. Although they do not belong to the definitional core of text and data mining, they may nonetheless render the exception intrinsically inapplicable. They are therefore examined prior to the other requirements of Article 3 and 4 CDSM Directive. Those other conditions concern the access to and the use of the relevant data and are thus external to the structure of the process, whereas memorisation and the three-step test affect the intrinsic compatibility of G-AI training with the rationale and limits of the exception.

a) Memorization

[94] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, p. 53.

Both Articles 3 and 4 require the reproduction not to be retained more than necessary for, on the one hand, the objective of the scientific research, and, on the other hand, to perform text and data mining. This difference is significant: under Article 3 it is possible to store reproductions and extractions for as long as it is justifiable by the research scope and verification, whereas Article 4, which may apply to any purpose, is considerably more restrictive. For this reason, the considerations below only apply to the latter.

As examined in §2.1, the question whether a trained AI model contains reproductions of its training material remains contested. This question has particular significance: if the model is regarded as containing reproductions of the training material, the retention limit would be systematically exceeded, since the model — by definition — persists beyond the completion of the training process. On this reading, the TDM exception would be structurally inapplicable to G-AI training under Article 4.

A further argument, as anticipated in §2.1.1., of a systematic nature that was supported in relation to the non-reproduction nature of the G-AI model, developed by Leistner and Antoine, supports the contrary conclusion. If trained AI models were to be regarded as reproductions of all the copyright-protected material used during training, this would sit uncomfortably with the structure of the TDM exception framework as a whole. The application of Articles 3 and 4 CDSM to AI training activities — together with the mere requirement to comply with and support the opt-out mechanism — would make no sense if the resulting trained model were itself a reproduction of the training material: the exception would authorise the training process only to prohibit its result. This internal contradiction is confirmed by the systematic connection between Article 4 CDSM and Article 53 AI Act. The latter obliges GPAI providers to put in place a policy to comply with EU copyright law and to identify and respect opt-outs pursuant to Article 4(3) CDSM — an obligation that presupposes that AI training falls within the scope of Article 4 and that the resulting model is not itself a reproduction of the training

material. It can therefore be derived from the EU legislator's design of the AI Act that the theory of trained AI models containing reproductions of their training material is not the approach the legislator intended to adopt.[95]

b) Compliance three-step test

Article 7 of the CDSM imposes an additional element to the TDM exception outlined in Articles 3 and 4, stipulating that these must be in accordance with Article 5(5) of the Infosoc Directive. As stated above, the latter is a transposition of the three-step test, which is a condition present in various international agreements. The EU is bound by the three-step test through the TRIPS Agreement and the WCT (*supra*, §1.1.) The following three conditions must be met for an exception or limitation of copyright to be lawful: (i) they apply only to certain special cases; (ii) they do not conflict with the normal exploitation of the work; (iii) and they do not unreasonably prejudice the legitimate interests of the author.

In order to address this provision, it is first necessary to establish precisely who is meant by the actual addressee of both the international and the EU directive provisions — namely, whether the obligation refers to regional and national legislators, or whether it also constitutes an interpretative criterion for national courts.

Under international copyright law, there is common agreement that the three-step test constitutes an obligation addressed exclusively to national legislators when drafting exceptions and limitations, and does not furnish grounds which can be raised in the course of private litigation.[96] If this interpretation were to be extended to Article 5(5) of the InfoSoc Directive and, by virtue of Article 7 CDSM — which operates as a cross-reference incorporating the full force of the InfoSoc three-step test into the TDM exception framework — to Articles 3 and 4 CDSM, it would

[95] Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §III.2. https://doi.org/10.1093/grurint/ikaf114

[96] Arnold, R., & Rosati, E. (2015). Are national courts the addressees of the InfoSoc three-step test? Journal of Intellectual Property Law & Practice, 10(10), 741, p. 742.

mean that national courts would only be required to interpret national exceptions in the light of EU law to the extent that national law is ambiguous, without being directly bound by the three-step test in individual cases.[97]

However, this interpretation is difficult to reconcile with the CJEU's case law on Article 5(5) of the InfoSoc Directive. Although the question has not been definitively resolved, the Court's jurisprudence provides clear indications that the three-step test is addressed at national courts as well as national legislators.[98] More specifically, the CJEU has held, on the one hand, that Article 5(5) is not intended to affect the substantive content of the exceptions and limitations contained in Article 5(1)–(4), and hence cannot extend their scope,[99] and that if acts clearly fall within one of those exceptions, they satisfy Article 5(5).[100] On the other hand, the Court has also held that exceptions and limitations must be construed in the light of Article 5(5),[101] and — more significantly — that despite the fact that Article 5(5) takes effect only at the time that exceptions and limitations are applied by the Member States, the acts of the defendant in question must nonetheless satisfy the requirements of Article 5(5).[102]

The consequence of this reading is significant: even in those Member States that have not transposed the language of the three-step test into their own legal systems, national courts must determine not only whether the acts of the defendant are eligible for the application of a given exception or limitation, but also whether they comply with the cumulative conditions set in the three-step test.[103] This may in

[97] Geiger, C. (2007). From Berne to national law, via the Copyright Directive: the dangerous mutations of the three-step test. European Intellectual Property Review, 29(12), 486, p. 488; Arnold & Rosati (2015), p. 742.
[98] Arnold, R., & Rosati, E. (2015), pp. 744–747; Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press, p. 56.
[99] CJEU, Spiegel Online, C-516/17, para. 79; ACI Adam, C-435/12, para. 26; Ulmer, C-117/13, para. 47; Copydan Båndkopi, C-463/12, para. 90.
[100] CJEU, Infopaq II, C-302/10, paras. 55–57.
[101] CJEU, Infopaq I, C-5/08, para. 58; OSA, C-351/12, para. 40; Ulmer, C-117/13, para. 47; Filmspeler, C-527/15, para. 63.
[102] CJEU, Football Association Premier League, C-403/08 and C-429/08, paras. 180–181; Public Relations Consultants Association, C-360/13, paras. 53–63; Ulmer, C-117/13, paras. 52 and 56; Filmspeler, C-527/15, paras. 65–66 and 69–71.
[103] Arnold, R., & Rosati, E. (2015), p. 747.

some instances result in the further restriction of available national exceptions and limitations, requiring national courts to disapply them in whole or in part where their application would breach Article 5(5) in the particular case at hand.[104]

This conclusion is particularly relevant in the context of the CDSM Directive. As noted above, Article 7 CDSM expressly requires the TDM exceptions to comply with Article 5(5) of the InfoSoc Directive. As legal scholarship has observed, compliance with the three-step test appears to be mandated upon both national legislatures and courts, and the mandatory character of the TDM exception in Article 3 and the broader exception in Article 4 must be applied in such a way that their effectiveness — including their underlying rationale — is safeguarded.[105]

In view of the recognised function of the three-step test in the interpretation of exceptions and limitations, it is essential to assign a precise meaning to the three requirements, and to assess whether G-AI training complies with them. The jurisprudence of the CJEU is not particularly extensive on the topic. However, it is possible to rely on the international interpretation, particularly as it is the original source of the provision. The most relevant document in this case is the WTO Panel Report, US – Section 110(5), which was a consultation requested by the European Communities in relation to Section 110(5) of the US Copyright Act.[106] It provides the interpretation to be given to Article 13 of the TRIPS Agreement. Firstly, with "in certain special cases", the report interpretation requires "clearly defined and narrow in quantitative as well as a qualitative sense". Secondly, the concept of "no conflict with normal exploitation" is explained, with the assertion that the economic income or potential income must not be significant. Thirdly, the principle of 'no unreasonable prejudice to legitimate interests' signifies that the

[104] Ibidem, p. 745; CJEU, OSA, C-351/12, Opinion of AG Sharpston, paras. 38, 43 and 46.
[105] Rosati, E. (2021). Copyright in the Digital Single Market. Oxford University Press, pp. 56–57 and 89. Furthermore, in relation to the interpretative character of Article (5) InfoSoc, see M. Bertani, Diritto d'autore europeo, Giappichelli, Torino, 2011 (Quaderni di AIDA, 21), cap. IV, §8.
[106] WTO Panel Report, United States — Section 110(5) of the US Copyright Act, WT/DS160/R, 15 June 2000. Section 110(5) of the US Copyright Act exempts certain establishments from paying royalties for the broadcast of music.

usage in question must not result in, or have the potential to result in, an unreasonable loss of income to the rightholder.

It has been asserted by certain legal scholars that the conditions aforementioned are not met in the case of TDM under Article 4 CDSM applied to the G-AI training. It is important to note that there are different degrees of objection.

In particular, certain scholars argue that there is a complete incompatibility between the TDM of Article 4 (Article 3 is not addressed since the issues might arise from the commercial exploitation) applied to G-AI training and the three-step test. It has been asserted that this would constitute a violation of all three conditions stipulated in Article 5(5) of the Infosoc Directive. With regard to the first condition, it is argued that the G-AI training method is deemed to be incompatible due to its requirement of extensive content intake. This requirement would be incompatible because, as we have seen above from the meaning attributed by the Panel, the speciality must be narrow also from a quantitative point of view. Secondly, as previously discussed in the context of memorisation, the G-AI model is, by some scholars, reputed to be capable of replicating training material with a high degree of precision (as also above analysed, advancements in this field have been noted), resulting in a loss of income for the author. Additionally, even when the output is not an exact or near-exact copy, it has been asserted that the model has the potential to generate content that rivals copyrighted material used for the training. Thirdly, the author's interests are not aligned, as the use of their work is both opaque and unlicensed.[107]

Conversely, alternative literature perspectives adopt a less critical stance, emphasising a distinction between the model's training and its practical commercial availability and utilisation. In particular, the former is not generally regarded as problematic, since it does not entail the possibility of any conflict with the interests of the rightholder. It is noteworthy that, in the context of the

[107] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs.

aforementioned position, the absence of publication would not preclude issues pertaining to the "certain special cases" condition. However, according to the present scholars, the consumption of substantial quantities of content does not appear to be at odds with the "special cases" requirement. In any case, with the programme's public availability, a potential conflict with the author's work emerges as a concern in relation to the second and third requirements of the three-step test.[108]

A further and analytically distinct argument has been advanced by Leistner and Antoine, who contest the incompatibility thesis on the basis of the causal relationship between training and output. Their starting point is that the link between the use of a copyrighted work for training purposes and an AI-generated output that competes with that work is far too vague and depends on a variety of circumstances which cannot be predicted with sufficient certainty, so that a conflict with the normal exploitation of the work cannot be assumed on this basis.[109] In addition, given the existence and acceptance of fair use in US copyright law, it would seem far-fetched to assume a violation of the three-step test simply because of the application of Article 4 CDSM to AI training activities.[110] A comparative perspective is offered by the Japanese framework, which contains an explicit restriction for cases in which exploitation would unreasonably prejudice the interests of the copyright owner. According to the Japanese Agency for Cultural Affairs, only two specific case groups may fall under this qualification: databases for TDM activities which contain full copies of journalistic contributions, thereby constituting a news database; and constellations in which a TDM-trained AI model massively produces content similar to the style of an individual artist, whose works

[108] European Copyright Society. (2025). Copyright and generative AI: Opinion of the European Copyright Society.
[109] Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §III.3. https://doi.org/10.1093/grurint/ikaf114
[110] Ibidem.

are largely replaced in the marketplace as a result.[111] These examples illustrate that situations in which rightholders' legitimate interests are unreasonably prejudiced only exist, if at all, in highly specific and particular case groups, and that an unreasonable prejudice cannot be generally assumed in respect of AI training activities under Article 4 CDSM.

In the context of the present discussion, the two German cases, LAION and GEMA, mention, directly or indirectly, the three-step test, but they are not particularly pertinent. It is evident that the former, despite referring to the test, does not address G-AI training directly[112], but the creation of a dataset (*supra*, §2.2.1.). The GEMA case does not explicitly mention the three-step test; however, it states that admitting the legitimacy of reproductions for G-AI training potentially compromises the right holders' enjoyment of their rights.[113] Nevertheless, as previously indicated in the context of memorisation, this issue may soon be resolved through technological innovations.

A significant element to be considered for the three-step test is the possibility under Article 4(3) CDSM to reserve the right, as it represents the balancing mechanism that the EU legislator has incorporated into the framework precisely in order to achieve a fair balance and comply with the test. As Leistner and Antoine observe, the opt-out mechanism gives rightholders the opportunity to prohibit the use of their works, thereby ensuring that the third element of the three-step test — the unreasonable prejudice to legitimate interests — is not violated as a general matter.[114] Nonetheless, it is the opinion of some scholars that it would be fallacious to assume that, in instances where a rightholder has not exercised this option, their

[111] Ibidem, §III.3, citing Japanese Agency for Cultural Affairs, General Understanding on AI and Copyright (2024).
[112] District Court of Hamburg (Landgericht Hamburg), judgment of 27 September 2024, 310 O 227/23, para. 84.
[113] GEMA v OpenAI, Landgericht München I, Az. 42 O 14139/24, 11 November 2025, para. 208.
[114] Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §III.3. https://doi.org/10.1093/grurint/ikaf114

works could be used for G-AI training purposes without any limitations.[115] Furthermore, as will be analysed in Chapter II §2.4. the opt-out might not be a theoretically reliable solution.

It is evident that in the case of Article 3 the balance is provided by the narrow scope and the limited entities that can perform it.

Whilst the doctrinal debate remains open, the European Parliament's resolution of 10 March 2026 on copyright and generative artificial intelligence expressly states that "any exceptions to current rights under the copyright acquis, including those relating to AI, must be in line with the three-step test [...]",[116] thereby confirming that any regulatory intervention in this area must comply with this criterion.

[115] Rosati, E. (2024). No step-free copyright exceptions: The role of the three-step in defining permitted uses of protected content (including TDM for AI-training purposes). European Intellectual Property Review, 46(5), 262–274.
[116] European Parliament, Report on copyright and generative artificial intelligence — opportunities and challenges, P10_TA(2026)0066, para. 3.

CHAPTER II – Applying the TDM Exception to G-AI Training: Requirements, Challenges, and Enforcement

Summary: 1. Overview; 2. Articles 3 and 4 CDSM; 2.1. Scope and subjects of Article 3; 2.2. Opt-out; 2.3. Lawful Access; 2.4. Interpretative Issues in Relation to Article 5(2) of the Berne Convention; 3. The AI Act and the Code of Practice; 4. Enforcement Challenges and Consequences of Non-Compliance; 5. Perspectives for Legislative Development; 5.1. Standardisation and Centralisation of the Opt-Out Mechanism; 5.2. Transparency, Traceability, and the Territorial Dimension; 5.3. Like Company v Google and Possible Developments

1. Overview

In the context of G-AI training, the lawful exercise of the TDM exception is subject to compliance with both CDSM and AI Act provisions. Articles 3 and 4 of the CDSM are characterised by the presence of a single, overarching threshold requirement, namely lawful access. However, these articles diverge significantly in terms of their subjective scope, purposive conditions, and the safeguards they incorporate. With regard to the realm of copyright, the primary objective of the AI Act and its associated documentation is to enhance the implementation of Article 3 and 4 within the context of GPAI. In particular, Article 53(1)(c) and (d) require the GPAI providers to adopt a policy that is compliant and safeguards copyright and to publish a detailed summary of the contents used in the training. Indeed, the enforcement of lawful access and opt-out requirements may prove challenging, since it may be hard, as also highlighted in the previous Chapter, to trace back to the content used for the training of a G-AI model. In accordance with Article 53 and the Code of Practices, which comprise a Chapter on Copyright, GPAI providers are obliged to retain materials, albeit in summary form. While this is a beneficial outcome, it does not resolve all the issues. Firstly, there are still many problems in the opt-out feasibility and operability, as also addressed by the European Parliament resolution of 10 March 2026. Secondly, it remains difficult to verify the accuracy of the disclosures made by GPAI providers. Connected to this latter challenge, it is noteworthy that while the AI Act addresses GPAI

providers, the supply chain leading to a G-AI model is usually extensive and involves multiple distinct entities (e.g. in the LAION case, the non-profit created a dataset for AI, but not the model) and it is usually performed across jurisdictions.

The purpose of this chapter is to evaluate the aforementioned issues and potential solutions, with particular attention to the European Parliament resolution of 10 March 2026. To this end, after briefly recalling the requirements of Articles 3 and 4 CDSM, the analysis turns to the challenges arising in their application to G-AI training in light of the AI Act obligations, before addressing the consequences of non-compliance

2. Articles 3 and 4 CDSM

As previously stated, Articles 3 and 4 of the CDSM differ in terms of the subjects and purpose of protection. Article 3 is applicable exclusively to research organisations and cultural heritage institutions for the purpose of scientific research. By contrast, Article 4 is relevant to any subject and for any purpose. The broader applicability of Article 4 is balanced through its third paragraph, which allows the reservation of the use of works or other subject matter by the rightholder (so-called opt-out). Conversely, this rightholder's safeguard would not impact Article 3, a matter that makes it especially significant to delineate the boundaries thereof. In any case,both provisions require lawful access to the content used during the performance.[117]

2.1. Scope and subjects of Article 3

[117] For a complete and systematic description of Articles 3 and 4 CDSM Directive see also P. Auteri, G. Floridia, V.M. Mangini, G. Olivieri, M. Ricolfi, R. Romano, G.E. Sironi, P. Spada, Diritto industriale. Proprietà intellettuale e concorrenza, 7ª ed., Giappichelli, Torino, 2023, Parte VII, cap. IV, §15.

The CDSM itself provides the definition of research organisation and cultural heritage institutions. The former is broader in scope, encompassing any entity that engages in scientific research for the purpose of advancing knowledge in a non-profit capacity, reinvesting in scientific research, or pursuing a public interest mission recognised by the Member State.[118] This category encompasses a variety of legal forms and structures. For instance, it may encompass both private and public entities, in addition to partnerships with private companies, provided that the influence of the latter on the research is not decisive.[119] With regard to the definition of cultural heritage institutions, the term is more limited in scope, encompassing exclusively "[...] publicly accessible library or museum, an archive or a film or audio heritage institution."[120]

As regards the practical application of the research organisation definition, this is illustrated by the LAION case, wherein the defendant (LAION) maintains its designation as a "research organisation" despite its collaboration with private companies and the concurrent employment of two of its employees by a partner company.[121] It should be noted that any subsequent use by a commercial organization of a dataset or model developed under Article 3, must be subject to a separate and distinct evaluation. In any case, this cooperation between non-profit and commercial companies can give rise to some challenges (*infra*, §3.)

Furthermore, it is worth specifying that the non-profit character required by Article 3 does not categorically exclude entrepreneurial actors from the definition of research organisation. The Italian transposition in Article 70-ter(4) LDA reproduces the two alternative qualifying conditions of the Directive: an entity qualifies either where it operates on a not-for-profit basis or by reinvesting its profits in scientific research, or where it pursues a public-interest mission

[118] 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, OJ L 130, 17.5.2019, pp. 92–125, Article 2(1).
[119] Ibidem, Recital 11.
[120] Ibidem, Article 2(3).
[121] District Court of Hamburg, 27.09.2024 - 310 O 227/23, paras. 114-116

recognised by a Member State. As Italian scholarship has observed, both conditions may in principle be satisfied by entrepreneurs, and the second may be met even by undertakings carrying on a purely commercial activity, provided they benefit from a public recognition of their research mission.[122]

Conversely, CDSM does not provide a definition for the purposive characteristic of Article 3. Indeed, with regard to the objective of scientific research, the Directive merely stipulates in Recital 12 that it may encompass both scientific and humanistic sciences. In any case, the definition may be analogous to that of TDM. Pursuant to the jurisprudence of the CJEU, it is necessary to consider the everyday language meaning of a term.[123] It can be inferred that this term encompasses activities aimed at generating new information and knowledge or insights that are based on or characterised by the methods and principles of science.[124] It may therefore be argued that the definition of scientific research does not inherently necessitate the absence of commercial elements.[125]

In summary, in the event of a TDM process aligning with the aforementioned subjective and purposive criteria, it will be entitled to the extensive copyright limitations stipulated in Article 3. In all other instances, the matter will be addressed within the parameters of Article 4.

2.2. Opt-out

[122] L. Mansani, 'Le eccezioni per estrazione di testo e dati, didattica e conservazione del patrimonio culturale', in AIDA, 2019; Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §13, pp. 1180–1181, drawing on Calabrese, B., Sulla dimensione imprenditoriale e societaria degli organismi di ricerca a fini di text and data mining, Contratto e impresa, 2020.
[123]CJEU, Constantin Film Verleih GmbH v YouTube LLC and Google LLC, C-264/19, EU:C:2020:542, para. 29.
[124] Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press.
[125] Ibidem.

In order to balance the broad scope of the Article 4 exception, which applies regardless of the purpose of the TDM activity, the EU legislator has provided rightholders with the possibility to reserve their works from TDM use. Article 4(3) stipulates that such reservation must be made expressly, in an appropriate manner, and by the rightholder. The first — the "express" requirement — primarily excludes implied declarations and requires that both the covered use acts and the content be identifiable with reasonable certainty. The second — the "appropriate manner" requirement — raises questions about the valid form of the opt-out, in particular as regards machine readability for online content, which is examined below. The third — the "by the rightholder" requirement — raises questions about who is entitled to declare a valid opt-out where rights are held or managed by multiple entities, also examined below.

In relation to the requirement of expressiveness of the reservation it is noteworthy that an opt-out is sufficiently specific if it allows the clear identification of the use acts covered — namely TDM and AI training activities — and of the content that is subject to the reservation. The Regional Court of Hamburg established, also supported by Leinster and Antoine, that the crucial factor is the possibility to identify the opt-out with reasonable certainty. In this sense, it is sufficient to have a single declaration covering all the works of a website, without the need for multiple reservations for each work.[126] This interpretation ensures a more workable mechanism in particular in the context of G-AI training, since it implies the ingestion of a vast quantity of material a case-by-case analysis would be difficult to impossible. [127]

When addressing the G-AI training the complexities and inefficiencies that inherently arise from the TDM exceptions are even accentuated. Indeed, it should

[126] District Court of Hamburg (Landgericht Hamburg), judgment of 27 September 2024, 310 O 227/23, para. 98; Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044, §III.4.a. https://doi.org/10.1093/grurint/ikaf114

[127] Ibidem.

be noted that the training content needed in the context of G-AI is different from the one generally used for traditional TDM. The former necessitates high quality expressive works, while the latter is mainly concerned with factual or structured data. Noticeably, the better the training material the more sophisticated the generative capabilities will be.[128] This is also a matter of social impact of G-AI systems, as also acknowledged by the European Parliament, since it ensures that the outputs they generate are reliable and of sufficient quality.[129] It is especially because of this structural dependency that the opt-out operability is so important in the field of G-AI training. Namely it is more probable that the most valuable content for the training is also the one most likely to be subject to reservation.

Furthermore, the massive scale of content ingestion required for G-AI training has made them more acute. Four problems may be identified: the uncertainty surrounding the machine readability requirement, the fragmentation of opt-out methods, the localisation of the reservation, and the overlapping of rights held by different rightholders.

Preliminarily, an important challenge concerns the concept of machine readability contained in Recital 18 CDSM in relation to content made publicly available online. It requires opt-outs to be expressed through machine-readable means, and it refers to metadata and website terms and conditions as examples. In any case, neither the CDSM nor its implementing legislation provides a precise definition of what constitutes machine readability in this context, and there are different positions in the scholarship.

There are two competing approaches that emerged both in the literature and in case law. The first and most narrow interpretation suggests that machine readability only encompasses reservations encoded in a structured, automatically processable format (e.g. a specific metadata tag or a standardized protocol), making it possible

[128] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, p. 56.
[129] European Parliament resolution of 10 March 2026 on copyright and generative artificial intelligence, P10_TA(2026)0066, recital V.

to identify them though an automated system without human intervention. In this perspective, opt-outs that are expressed in natural language, for example general website terms and conditions prohibiting web scraping, would not satisfy the machine readability criterion.[130]

A broader interpretation, which finds support in the Hamburg court's obiter dicta in the LAION case, holds that reservations expressed in natural language may satisfy the machine readability criterion, on the basis that AI developers are required by Article 53(1)(c) of the AI Act to identify and comply with rights reservations through state-of-the-art technologies, which include natural language processing capabilities capable of reading and interpreting website terms and conditions.[131] On this reading, the standard of machine readability is not static but evolves with the technological capabilities of TDM users — a dynamic interpretation that is consistent with the AI Act's reference to state-of-the-art technologies but raises significant legal certainty concerns.

The broader interpretation, while arguably more protective of rightholders, generates a further asymmetry: if the machine readability standard is modulated on the technological capabilities of the actor performing the TDM activity, the same reservation may be valid against an AI developer subject to the AI Act obligations but invalid against a different TDM user who is not. This would create an inconsistent enforcement landscape in which the validity of an opt-out depends not only on how it is expressed but also on who reads it — an outcome that appears

[130] Hamann, H. (2024). Artificial intelligence and the law of machine-readability: A review of human-to-machine communication protocols and their (in)compatibility with Article 4(3) of the Copyright DSM Directive. JIPITEC — Journal of Intellectual Property, Information Technology and E-Commerce Law, 15(2). In support of this narrow interpretation, Hamann refers to the definition of "machine-readable format" in Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, OJ L 172, 26.6.2019, pp. 56–83, which defines machine-readable format as "a file format structured so that software applications can easily identify, recognise and extract specific data." On this basis, Hamann argues that a reservation expressed solely in natural language does not meet the machine readability criterion, since it is not structured in a way that allows for automatic identification and extraction by software.

[131] District Court of Hamburg, 27.09.2024, 310 O 227/23, para. 103.

difficult to reconcile with the objective character of the lawful access requirement (*infra*, Chapter II, §2.3), and with the principle of legal certainty.[132]

As regards fragmentation, Recital 18 CDSM provides that, in the case of content made publicly available online, the reservation must be expressed by machine-readable means, including metadata and terms and conditions of a website or service; for other types of content, contractual agreements or unilateral declarations suffice, provided they are accompanied by effective technological measures within the meaning of Article 6(1) and (3) of the InfoSoc Directive.[133] Since no single opt-out mechanism has emerged as the prevailing standard, legally-driven and technical measures currently coexist.[134] The legally-driven measures include unilateral declarations, licensing constraints, and website terms and conditions, while the technical measures encompass a range of protocols and standards: the Robots Exclusion Protocol (hereinafter, REP), the TDM Reservation Protocol (TDMRep), robots meta tags, and asset-based standards such as the C2PA Content Authenticity Initiative and the JPEG Trust standard, in addition to proprietary services developed by specialised providers.[135]

The Robots Exclusion Protocol currently serves as the de facto standard for managing web crawling and scraping activities and has been widely deployed as the primary strategy for TDM rights reservations. However, as the EUIPO has noted, there is a prevailing consensus among stakeholders that REP is not optimal as a TDM opt-out mechanism and functions more as a temporary solution. Its principal limitations include its inherently limited granularity and use-specificity, its dependence on intermediation by website managers, its unenforceability, and

[132] EUIPO. (2025), p. 75; Lucchi, N., & Hunter, S. (2025). Generative AI and copyright – Training, creation, regulation. European Parliament, PE 774.095, p. 58.
[133] On the notion of effective technological measures within the meaning of Article 6(1) and (3) of the InfoSoc Directive, see CJEU, Case C-392/19, VG Bild-Kunst v Stiftung Preußischer Kulturbesitz, ECLI:EU:C:2021:181.
[134] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 218.
[135] Ibidem, pp. 218-220.

its reliance on the voluntary disclosure of web crawler identities.[136] In particular, REP operates at the level of the website or URL, and cannot be configured to reserve rights in relation to specific works or specific uses within a broader corpus of content. This limitation is particularly significant in the context of G-AI training, where the relevant content may be dispersed across thousands of different locations and platforms, each requiring a separate reservation.

The coexistence of multiple legally-driven and technical measures does not provide more flexibility; instead, it generates significant compliance uncertainty for both rightholders and AI developers. On the one hand, the fragmented framework makes it challenging for rightholders to navigate thought the different measures that may be required for different types of content, different platforms, and different jurisdictions, even with the risk that a reservation expressed through one mechanism will not be recognised as valid under the standards applied by a particular AI developer. On the other hand, AI developers must respect and be capable of processing different kinds of signals across many protocols and formats, a task that increases in complexity in consideration of the scale of the training data.[137]

Thirdly, one issue might arise from the localisation of the opt-out. Indeed, location-based measures, such as REP, operate by associating the reservation with a specific URL or website, and they are not inherent in the work. Namely, a rights reservation expressed through robots.txt on a particular website does not automatically extend to other locations where the same work may be lawfully accessible (e.g. where the same content has been licensed to multiple platforms or aggregators). As the EUIPO has observed, a copyright owner who wishes to universally opt out of TDM may therefore need to coordinate the expression of the reservation with all licensees and platforms through which its content is made available. Evidently, the

[136] Ibidem, p. 200.
[137] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, p. 58.

latter is a task that is onerous where content is distributed through complex multi-party licensing chains.[138]

The situation worsens when both asset-based and location-based techniques are involved. An asset-based measure indicates a signal associated with a specific digital file (e.g. a metadata), and it may indicate that the rightholder has not reserved TDM rights, while the website hosting that file may have expressed a location-based opt-out through its terms and conditions or REP instructions. On the contrary, it is also possible that a file carries embedded metadata reserving TDM rights, while being hosted on a platform whose terms of service do not restrict web scraping. In such cases, AI developers face genuine uncertainty as to which reservation signal should prevail. This might bring the AI provider to ignore a valid reservation or exclude content that the underlying rightholder had not in fact intended to reserve.[139]

The fourth problem arises from the structure of copyright and related rights, which in many cases vest in multiple rightholders whose positions on TDM reservation may diverge. Article 4(3) CDSM requires the opt-out to be made "by the rightholder", a requirement that is straightforward where a single author holds all relevant rights but becomes significantly more complex in the context of works subject to multi-layered rights structures.

Different exclusive rights within the bundle conferred by EU copyright law may be held or managed by different parties. In the musical sector, for instance, performance rights and mechanical rights — the latter being the reproduction right that is directly implicated by Article 4 — may be managed by different collective management organisations. Since Article 4 is an exception to the reproduction right specifically, it may be the party managing that right in particular, rather than the CMO representing communication to the public rights, that is the relevant party

[138] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, pp. 74–75.
[139] Ibidem, p. 232.

to make a valid opt-out.[140] This distinction is practically significant: a CMO that has declared an opt-out on behalf of its members — as GEMA did following amendments to its membership authorisation agreement in 2022 — must ensure that its mandate specifically extends to the right of reproduction, failing which the reservation may be ineffective.[141]

Furthermore, another layer of complexity arises with the involvement of commercial intermediaries or aggregators. In relation to that, the LAION case confirmed in obiter dictum that a licensee may in principle make a valid TDM reservation on behalf of the copyright owner, on the basis of agency principles derived from the licensing relationship.[142] However, this does not solve the problem completely: when there is more than one licensee, or where the licence is non-exclusive and the content is distributed through multiple platforms, the question of which party is authorised to make a reservation (and whether a reservation made by one licensee binds the others) remains largely unresolved. It should also be recalled that this challenge is even worse considering the multi-jurisdictions dimension of certain works, which may have different rightholders in different places: meaning that a reservation expressed in one Member State may not automatically extend to uses occurring in others.[143]

In the audiovisual sector these challenges are even more acute, as the contractual arrangements are usually more complicated, entailing many different rights (neighbouring rights and copyright) which increase the overlapping. Namely, these circumstances have been identified as obstacles for the licensing markets for G-AI training.[144]

The inadequacy of the current opt-out system in relation to G-AI training has been further addressed at the EU level, as will be discussed in the following sections

[140] Ibidem, p. 46.
[141] Ibidem, p. 47.
[142] District Court of Hamburg, 27.09.2024, 310 O 227/23, para. 99; EUIPO. (2025), p. 47.
[143] EUIPO. (2025), p. 47.
[144] Ibidem, p. 97.

with reference to the obligations introduced by the AI Act and the relevant provisions of the Code of Practice.

2.3. Lawful Access

A common characteristic of Articles 3 and 4 CDSM is the requirement of lawful access. To understand the meaning of this condition, Recitals 14 and 18 must be considered. Recital 14 identifies three cases in which content is accessed lawfully: when it is available under an open access policy; when there is a licensing agreement; and when it is freely available online. Recital 18 further underscores the relevance of lawful access in the context of Article 4, where the reservation of rights operates as an additional element in determining whether use of the content is compliant.

The condition was primarily designed to prevent the use of pirated content. In interpreting the lawful access requirement under Articles 3 and 4 CDSM, it is necessary to consider the CJEU's existing case law on the lawfulness of the source in the context of the reproduction right. In ACI Adam, decided in the context of the private copying exception under Article 5(2)(b) of the InfoSoc Directive, the Court held that the lawfulness of the source from which a reproduction is made is an objective condition for the application of that exception: a copy made from an unlawful source cannot benefit from the private copying exception, regardless of whether the user was aware of the unlawfulness of the source.[145] This objective approach is consistent with the nature of the reproduction right and with the strict interpretation required for exceptions thereto.

In relation to the right of communication to the public the CJEU's jurisprudence has identified a different logic. Indeed, in GS Media, the Court introduced a subjective element, grounded in the presumption of knowledge of unlawfulness and modulated on the profit-making nature of the activity. This means that entities

[145]CJEU, ACI Adam BV and Others v Stichting de Thuiskopie and Stichting Onderhandelingen Thuiskopie vergoeding, C-435/12, EU:C:2014:254, para. 41.

acting for profit are presumed to know that the content they exploit has been posted without the rightholder's consent, whereas private users benefit from a presumption of ignorance.[146] However, this distinction is specific to the making available right and is linked to the communicative nature of the act, rather than to the act of reproduction as such. Therefore, it does not translate to the reproduction right, which governs the acts of copying and storage involved in TDM.[147]

In this sense, the approach developed in ACI Adams appears more appropriate for the purposes of Articles 3 and 4 CDSM. In this view, the lawful access is to be interpreted objectively and without reference to the subjective nature of the entity performing TDM.[148] As a consequence, the content accessed in violation of technological protection measures, contractual restrictions, or rights reservations does not respect the lawful access, regardless of whether the TDM operator was aware of such restrictions.

Nevertheless, certain interpretative challenges arise, the requirement is not uniformly transposed across Member States: in Slovenia, for instance, freely accessible content is not included within the scope of lawful access, in contrast with what Recital 14 appears to contemplate.[149] This gives rise to concerns regarding the coherence of the internal market.[150]

---

[146] CJEU, GS Media BV v Sanoma Media Netherlands BV and Others, C-160/15, EU:C:2016:644, paras. 49–51.

[147] A divergent interpretation has been proposed in particular by Margoni, who argues on two grounds: first, that the concept of "lawful source" should be aligned with the subjective approach developed in GS Media, differentiating between profit-making entities — subject to a presumption of knowledge of unlawfulness — and non-profit entities such as research organisations, which would benefit from a presumption of ignorance; second, that even where a research organisation becomes aware of the unlawfulness of a source, the obligation to delete the relevant content may prove particularly burdensome and costly in the context of trained AI models, effectively rendering the Article 3 exception ineffective. As an alternative, the author proposes admitting all content freely accessible on the public internet — with the exception of content protected by a paywall or other technological measure — as lawfully accessible for the purposes of Article 3, in light of its scientific, cultural, and economic importance. See Margoni, T. (2024). TDM and generative AI: Lawful access and opt-outs. Forthcoming in Auteurs & Media 2024. SSRN. https://ssrn.com/abstract=5036164

[148] Noticeably, this would create a difference in the application of Article 3 and 4 CDSM.

[149] Bogataj Jančič, M., & Purkart, E. (2025). Text and data mining in the Slovenian legal system. Stockholm Intellectual Property Law Review, 2024(2), 5–8.

[150] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament.

The AI Act and the Code of Practice for GPAI have introduced transparency obligations that are relevant to the verification of lawful access compliance, in particular as regards the content used for model training. These provisions will be examined in the following sections.

2.4. Interpretative Issues in Relation to Article 5(2) of the Berne Convention

A distinct question, logically prior to the practical difficulties examined above, concerns the legitimacy of the reservation mechanism itself in light of the prohibition of constitutive formalities under Article 5(2) of the Berne Convention, pursuant to which the enjoyment and the exercise of copyright may not be subject to any formality. Examining the analogous reservation under Article 70-quater of the Italian Copyright Act, Servanzi subjects to detailed scrutiny the question whether the opt-out — a positive act required of the rightholder in order to retain control over the use of the work for data mining — amounts to a prohibited formality as such: he surveys a series of arguments advanced in support of its legitimacy and finds them, in the main, unpersuasive. Central to his analysis is the identification of a distinct faculty: alongside reproduction and extraction, a further power to control the data mining of the work, conceived — on the continental understanding of copyright as a general *ius excludendi alios* rather than a closed list of enumerated economic rights — as an autonomous form of economic exploitation, and as such "relative to the work".[151] If that power is so characterised, conditioning it on the opt-out makes the reservation a constitutive formality as to that very faculty, with the consequence that rightholders could retain control over the mining of their works, and over the reproductions ancillary to it, even in the absence of any reservation.[152] If the reservation were so characterised, rightholders

[151] For the fuller account of this further faculty and of the general ius excludendi from which it derives, see supra note 30.

[152] The question is developed, in relation to Article 70-quater LDA, by Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §§4–7, pp. 1158–1165. The same logic applies to the reservation under Article 4(3) CDSM, of which Article 70-quater is the national transposition.

could retain control over the mining of their works, and over the reproductions ancillary to it, even in the absence of any reservation.[153] In practice this incidence is felt chiefly in relation to non-EU rightholders, against whom the obligations deriving from EU law are not opposable.[154]

It bears emphasis that the foregoing reasoning is confined to copyright proper. It does not extend to the sui generis database right (supra, Chapter I §2.1.2), for which no international instrument imposes a prohibition of constitutive formalities, so that neither the Union nor its Member States are bound to guarantee its protection independently of any formality; in respect of that right the legitimacy concern examined here does not arise[155]

3. The AI Act and Code of Practice Obligations

Focusing on the specificity of G-AI training, it is relevant to note that the intervention is not brought about through the CDSM Directive, which does not address the challenges of the fragmentation of the opt-out mechanism and the interpretative difficulties surrounding lawful access. Conversely, the EU legislator sought to mitigate these difficulties through the transparency obligations introduced by the AI Act. These obligations, despite operating on a distinct regulatory plane, interact significantly with the copyright framework.

In Article 53(1) AI Act, two specific obligations are imposed on GPAI providers, which are relevant in this context. Firstly, Article 53(1)(c) stipulates that providers must implement a policy in order to comply with Union law on copyright and related rights. This policy must include the identification of rights reservations

[153] Ibidem, §4, pp. 1158–1159.
[154] Servanzi notes that one of the functions of Article 5(2) is to prevent foreign authors from being placed in a position less favourable than nationals, an effect the reservation system risks producing where a foreign rightholder, unaware of the need to reserve, omits to do so and is thereby subjected to the exception. Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §4, p. 1164.
[155] Ibidem, §11, pp. 1174–1175, where the author notes that the analysis is relative to copyright alone and that the extension to related rights — and in particular to the non-creative database right, which is not the object of any international treaty imposing the prohibition of formalities — would require separate examination.

expressed pursuant to Article 4(3) CDSM, and compliance with these reservations, including through state-of-the-art technologies. Secondly, Article 53(1)(d) stipulates that providers must prepare and make publicly available a sufficiently detailed summary of the content used for training their models, in accordance with a template provided by the AI Office. The underlying rationale for these obligations is articulated in Recitals 105 to 108, which explicitly acknowledge that TDM techniques may be used extensively in G-AI training in relation to copyright-protected content, and that compliance with the opt-out mechanism is a prerequisite for the lawful exercise of the Article 4 exception.[156]

Such obligations are further elaborated in the Code of Practice for General-Purpose AI Models, whose Copyright Chapter entered into force on 2 August 2025. It should be noted that adherence to the Code is entirely voluntary, and that there is no conclusive evidence that this indicates compliance with the obligations set out in the AI Act. [157] Nevertheless, it proposes a more comprehensive operational framework for demonstrating compliance with Article 53(1)(c). In particular, Commitment 1 stipulates five measures: the adoption, maintenance, and publication of a copyright policy (Measure 1.1); the obligation to reproduce and extract only content that is lawfully accessible when crawling the web, including by not circumventing effective technological measures and by excluding websites that persistently and repeatedly infringe copyright on a commercial scale (Measure 1.2); the identification and compliance with machine-readable rights reservations expressed pursuant to Article 4(3) CDSM, including through the Robot Exclusion Protocol and other appropriate protocols (Measure 1.3); the implementation of technical safeguards to mitigate the risk of copyright-infringing outputs (Measure 1.4); and the designation of a point of contact and complaint mechanism for affected rightholders (Measure 1.5).[158]

[156] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act), OJ L, 12.7.2024, Recitals 105–108.
[157]Code of Practice for General-Purpose AI Models, Copyright Chapter (2025), Objectives, point A.
[158] Ibidem, Commitment 1, Measures 1.1–1.5.

Measure 1.2 is particularly significant in the context of lawful access, as it directly addresses the risk of web-crawling from infringing sources, thereby complementing the lawful source requirement identified in the previous section. Measure 1.3 operationalises the obligation to identify and comply with rights reservations during web crawling. It does so through a two-tier commitment. The first tier requires signatories to employ web crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol, as specified in IETF Request for Comments No. 9309, including any subsequent version that is demonstrated to be technically feasible and implementable. The second tier requires signatories to identify and comply with other appropriate machine-readable protocols — for example through asset-based or location-based metadata — that have either been adopted by international or European standardisation organisations, or are state-of-the-art, technically implementable, and widely adopted by rightholders across different cultural sectors, and generally agreed through an inclusive process involving rightholders, AI providers, and other relevant stakeholders.[159]

This second tier is of particular significance insofar as it acknowledges, albeit implicitly, that REP alone is insufficient as a universal standard for rights reservations. It thus follows that a plurality of complementary mechanisms is necessary to ensure adequate coverage across different content types and distribution channels. Concurrently, the formulation — which conditions compliance on the existence of standards that are "widely adopted" and "generally agreed" — introduces a degree of circularity: the obligation to comply with a standard presupposes that such a standard exists, but the development of the standard itself depends on the voluntary engagement of the parties. In order to address this discrepancy, paragraph 3 of Measure 1.3 encourages signatories to participate in bona fide discussions with rightholders with a view to developing

[159] Ibidem, Measure 1.3(1).

appropriate machine-readable standards. It is important to note, however, that this commitment is explicitly framed as voluntary rather than binding.[160]

Paragraph 4 is also worthy of note, as it requires signatories to take appropriate measures to enable affected rightholders to obtain information about the web crawlers employed and their robots.txt configurations, and to provide a means for rightholders to be automatically notified when such information is updated. This transparency obligation operates alongside the right of information provided for in Article 8 of the Enforcement Directive, and is explicitly stated to be without prejudice to this right. This creates a complementary layer of disclosure that may facilitate rightholder enforcement without replacing the existing civil law remedies.[161]

However, as has been acknowledged by institutional, scholarly sources and main stakeholders,[162] these obligations do not resolve all of the challenges identified above. Three structural limitations are particularly significant in the context of G-AI training.

Firstly, the Code of Practice's provisions on rights reservation compliance pertain exclusively to data procured via web crawling, thereby excluding alternative prevalent data acquisition methodologies such as dataset downloads, API harvesting, or third-party aggregations.[163] The design choice in question is challenging to justify in the context of the broader mandate of Article 53(1)(c), which calls for a comprehensive copyright compliance policy irrespective of the manner in which the training data is procured.[164] There is a risk that this design

[160] Ibidem, Measure 1.3(3).
[161] Ibidem, Measure 1.3(4).
[162]EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office; Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament.
[163] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, p. 60.
[164] Ibidem

choice could inadvertently create incentives for providers to circumvent compliance by altering their data collection strategies.

Secondly, the supply chain leading to a G-AI model is typically extensive and involves multiple discrete entities, with no entity being responsible for the overall process in its entirety. As demonstrated by the LAION case, the entity responsible for creating the dataset and the entity responsible for training the model are frequently distinct. A further related concern pertains to the so-called 'data laundering' risk, whereby datasets originally compiled under the provisions of Article 3 of the Scientific Research Exception are subsequently reused in the context of commercial G-AI training under Article 4. This process potentially circumvents the opt-out mechanism altogether.[165] The AI Act exclusively addresses GPAI providers, thereby leaving this upstream fragmentation largely unresolved.

Third, it remains particularly difficult to verify the accuracy and completeness of the disclosures made by GPAI providers pursuant to Article 53(1)(d). The summary obligation is conceived in general terms and does not require a work-by-work assessment of the training data.[166] In this field, a further structural limitation concerns the scope of the supervisory role assigned to the AI Office under Recital 108 of the AI Act. The Recital explicitly provides that the AI Office shall monitor whether providers have fulfilled their transparency obligations without verifying or proceeding to a work-by-work assessment of the training data in terms of copyright compliance. This reflects a deliberate legislative choice to confine the AI Office's role to a formal assessment of whether providers have adopted a copyright policy and published a training data summary, without extending to the substantive accuracy or completeness of those disclosures. Such a choice is

[165] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 117. This risk was already identified in realtion to traditional TDM, see L. Mansani, 'Le eccezioni per estrazione di testo e dati, didattica e conservazione del patrimonio culturale', in AIDA, 2019, §8.
[166] AI Act, Recital 107; Lucchi, N., & Hunter, S. (2025), p. 59.

consistent with the broader regulatory design of the AI Act, which preserves the competence of national courts and the CJEU as the primary forums for copyright enforcement, as confirmed by Recital (b) of the Code of Practice.[167]

The practical consequence of this design, however, is that the substantive verification of whether specific copyright-protected works were used in training without authorisation or in violation of an opt-out reservation falls primarily on rightholders, who must independently establish this through private litigation. Whether this allocation of enforcement responsibilities is adequate in light of the structural information asymmetries identified above is a question that will be examined in the following section.

4. Enforcement Challenges and Consequences of Non-Compliance

The enforcement of copyright obligations in the context of G-AI training operates across two distinct but interrelated regulatory planes. On the one hand, the AI Act establishes an administrative-sanctioning regime: pursuant to Article 99, violations of the obligations imposed on GPAI providers under Article 53, including the transparency obligations examined in the previous section, are subject to administrative fines of up to 15 million euros or three percent of global annual turnover, whichever is higher.[168] On the other hand, the failure to comply with the conditions for the lawful exercise of the TDM exception — in particular the requirements of lawful access and opt-out — might constitute a violation of the exclusive rights conferred under EU copyright law and gives rise to civil liability under the framework established by Directive 2004/48/EC (hereinafter, Enforcement Directive), which provides for injunctive relief, provisional measures, and damages. It is also noteworthy that Member States can add further

[167]Code of Practice for General-Purpose AI Models, Copyright Chapter (2025), Recital (b): "This Chapter in no way affects the application and enforcement of Union law on copyright and related rights which is for the courts of Member States and ultimately the Court of Justice of the European Union to interpret."
[168] Regulation (EU) 2024/1689 (AI Act), Article 99(3).

forms of sanctions, for example of criminal law nature, and that the implementation of the Directive can differ among them.[169]

These two regimes are characterised by structural distinction and operate in an autonomous manner: a violation of Article 53 of the AI Act does not inherently result in a violation of copyright law. Conversely, copyright infringement is not necessarily indicative of a breach of AI Act obligations. Nevertheless, in actuality, the two are intimately associated. The failure of a GPAI provider to adopt a copyright compliance policy or to publish a sufficiently detailed training data summary may, on the one hand, prompt rights holders to investigate potential copyright infringements, since the absence or inadequacy of such disclosures may itself signal non-compliance with the conditions of the TDM exception; and, on the other, provide an additional basis on which rights holders can act, given that the AI Act administrative track offers a complementary avenue for addressing conduct that may simultaneously engage copyright concerns. The administrative sanction under the AI Act and the civil liability under copyright law thus can influence each other, even if they remain legally distinct.

Even though the violation of the AI Act provisions and the non-compliance with the TDM exception implies consequences, and this should help the enforceability, there are some challenges in the identification of those two conducts.

The first concerns the traceability of the supply chain from an enforcement perspective. Even assuming that a GPAI provider has published a training data summary in compliance with Article 53(1)(d), a rightholder seeking to establish copyright infringement faces a structural evidentiary obstacle: the information contained in that summary may not extend to the upstream phases of data acquisition and processing, which are typically carried out by entities — dataset aggregators, pre-training providers, and data brokers — that are not subject to any

[169] Member States may introduce additional sanctions of a criminal nature pursuant to Article 8(1) of Directive 2004/48/EC, which requires that measures, procedures and remedies be effective, proportionate and dissuasive. For Italy specifically, see Legge 22 aprile 1941, n. 633, Protezione del diritto d'autore e di altri diritti connessi al suo esercizio, Articles 171 et seq.

disclosure obligation under the AI Act. Reconstructing the full provenance of the content used for training may therefore require the rightholder to pursue information claims against multiple parties across different jurisdictions, a task that is both procedurally complex and practically onerous. The right of information under Article 8 of the Enforcement Directive, which in principle allows rightholders to obtain information about the origin and distribution channels of infringing content, may prove insufficient in this context, as the relevant information may be dispersed across entities that are not party to the copyright infringement proceedings.[170]

The second issue pertains to the verifiability of the declarations made by GPAI providers. As has been previously established, the overarching obligation stipulated under Article 53(1)(d) is delineated in broad terms and does not necessitate a comprehensive, work-by-work evaluation of the training data. According to Recital 108 of the AI Act, the responsibility of the AI Office does not extend to the verification of the accuracy or completeness of the published summaries. In circumstances where autonomous auditing mechanisms or technical standards for dataset documentation are absent, rights holders find themselves devoid of effective means by which to verify the utilization of their works for training purposes, or to ascertain whether the provider has adhered to any opt-out reservations that may be expressed. The aforementioned structural information inequality is further compounded by the scale of G-AI training, which involves billions of data points across multiple languages, formats, and jurisdictions.[171]

The third problem concerns the adequacy of the existing civil enforcement framework. The Enforcement Directive provides rightholders with a set of procedural instruments — including the right to obtain information about the origin and distribution channels of infringing content pursuant to Article 8, injunctive

[170] Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, p. 60.
[171] EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 228.

relief, and damages — that in principle apply to copyright infringements arising from G-AI training. However, as the Lucchi report has noted, this Directive was not conceived with the systemic opacity and industrial scale of G-AI training in mind, and its current tools may fall short in addressing the unique enforcement challenges posed by such training, particularly when conducted by non-EU providers or through decentralised supply chains.[172] The right of information under Article 8, in particular, may prove insufficient where the relevant data is dispersed across multiple entities in the supply chain, or where its disclosure is resisted on grounds of trade secrecy.

Besides these three structural challenges, the enforcement of the GPAI providers' obligations might be affected by the inconsistency and lack of coordination among different other pieces of legislation. Indeed, a further tension that has received insufficient attention in the literature is the conflict between the transparency obligations of the AI Act and the protection of confidential business information under Directive 2016/943/EU (Trade Secrets Directive). GPAI providers may legitimately oppose the disclosure of detailed information about their training datasets by invoking the protection of trade secrets, since the composition of those datasets may constitute commercially sensitive information. In the absence of clear procedural guidance on how to balance these competing interests, this tension risks undermining the enforceability of the Article 53 disclosures and creating a legal grey zone in which providers can invoke trade secrecy to limit the scope of their transparency obligations.[173]

In light of these structural challenges, the European Parliament resolution of 10 March 2026 has proposed a significant corrective mechanism: namely, the introduction of a rebuttable presumption that any G-AI model placed on the EU market has utilised copyright-protected works for training, inference, or retrieval-

[172] Lucchi, N., & Hunter, S. (2025), p. 154.
[173] Ibidem, p. 155; Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets).

augmented generation, where the statutory transparency obligations have not been fully complied with.[174] The aforementioned presumption would result in a shift of the burden of proof from the rightholder to the provider. Currently, the rightholder bears the near-impossible task of demonstrating the specific utilisation of works in a particular training process. The presumption would require the provider to demonstrate compliance to disprove the presumption. The resolution further recommends that, in instances where a rightholder prevails on the basis of such a presumption or on the evidence presented, the AI provider should be held liable for all reasonable and proportionate legal costs.[175]

This proposal signifies a substantial progression in the discourse surrounding the enforcement of copyright regulations, as it acknowledges the inadequacy of the prevailing burden of proof framework in aligning with the intricacies of G-AI training. It asserts the necessity for a normative rectification to ensure the effective enforcement of copyright in this domain. The question of whether such a presumption would be compatible with the general principles of EU procedural law, and how it would interact with the existing framework of the Enforcement Directive, remain questions that will require further legislative and judicial elaboration.

5. Perspectives for Legislative Development

The analysis conducted in the preceding sections has revealed a series of structural gaps in the current EU framework for G-AI training. The opt-out mechanism suffers from fragmentation, localisation problems, and coordination failures; the transparency obligations introduced by the AI Act, while representing a meaningful step forward, are not accompanied by adequate verification mechanisms; and the enforcement framework lacks the procedural tools necessary to address the information asymmetries that characterise the relationship between

[174] European Parliament resolution of 10 March 2026 on copyright and generative artificial intelligence (P10_TA(2026)0066), recital AB and para. 24.
[175] Ibidem, para. 24.

rightholders and GPAI providers. These gaps have prompted significant institutional and scholarly reflection on the direction that legislative development should take.

5.1. Standardisation and Centralisation of the Opt-Out Mechanism

A first area where reform appears both necessary and feasible within the existing framework concerns the standardisation of opt-out mechanisms. As the analysis of the fragmentation problem has shown, the current coexistence of legally-driven and technical reservation methods — REP, TDMRep, C2PA, embedded metadata, website terms and conditions — creates compliance uncertainty for both rightholders and AI developers without any of these mechanisms providing adequate coverage across all content types and distribution channels. The absence of a universally recognised standard means that rightholders face the impossible task of expressing reservations through multiple incompatible protocols, while AI developers must integrate complex decision rules to navigate conflicting reservation signals.[176]

In this regard, the European Parliament resolution of 10 March 2026 has proposed entrusting the EUIPO with the role of trusted intermediary responsible for managing and cataloguing opt-out declarations in a limited number of standardised machine-readable formats. The participation in the mechanism should be straightforward and cost-effective, and should not invalidate previously expressed opt-outs.[177] The proposal outlined above would address the fragmentation and localisation issues identified through creation of a single, authoritative registry of rights reservations. This register would be available to AI developers as a resource to aid their compliance processes. It is also noted by the resolution that such a mechanism should be designed through consultation with rightholders, AI providers, and their representative organisations. This would ensure that the

[176]Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, p. 57.
[177] European Parliament resolution of 10 March 2026 (P10_TA(2026)0066), para. 10.

mechanism is not only technically implementable but also practically accessible to individual creators and small-scale rightholders, who currently lack the resources to navigate the existing opt-out landscape.[178]

A further consideration pertains to the relationship between the opt-out mechanism and the evolution of a licensing market for G-AI training content. In accordance with Article 4 CDSM, the TDM exception is available ipso facto to any person, unless the rights holder has expressly reserved their rights. This suggests that the opt-out mechanism is not merely a defensive tool; rather, it is the precondition for the emergence of a commercial market for AI training content. In the event that a rights holder is unable to effectively exercise the opt-out, this results in the exclusion of the free use of their content under the exception. Consequently, the rights holder will be devoid of contractual leverage to negotiate a licence and obtain remuneration. Conversely, where the opt-out is effectively exercised, the provider is no longer able to rely on the exception and must instead seek authorisation, thus creating a commercial space for licensing agreements. [179] The EUIPO has observed that the standardisation of opt-out mechanisms would therefore serve a dual purpose: improving the enforceability of rights reservations on the one hand, and creating the structural conditions for a more efficient and transparent market for AI training content on the other.

5.2. Transparency, Traceability, and the Territorial Dimension

A second area of reform concerns the transparency obligations imposed on GPAI providers and their territorial scope. As examined in the previous sections, the current framework limits disclosure obligations to GPAI providers, leaving upstream actors in the data supply chain outside the scope of any reporting requirement. The Lucchi report has proposed addressing this gap through the introduction of standardised dataset logs and traceability tools — such as

[178] Ibidem, para. 10.
[179]EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office, p. 200.

watermarking and fingerprinting — that would allow for end-to-end auditing of copyright-protected content throughout the training pipeline.[180] These tools would complement the summary obligation under Article 53(1)(d) by providing a more granular and verifiable record of the content used, and would facilitate the exercise of the right of information under Article 8 of the Enforcement Directive in the context of copyright infringement proceedings.

In addition to the aforementioned, the resolution of 10 March 2026 has also addressed the territorial dimension of transparency obligations, proposing that providers and deployers of G-AI models placed on the EU market should be required to ensure full transparency regarding all copyright-protected content used for training, inference, and retrieval-augmented generation, regardless of the jurisdiction in which the relevant copyright acts took place.[181] This proposal directly addresses the risk identified by both the EUIPO and the Lucchi report of non-EU providers circumventing compliance by conducting training activities in jurisdictions with more permissive copyright regimes, such as the United States and Japan.[182] It is argued that the extraterritorial application of EU transparency obligations would ensure a level playing field between EU and non-EU providers, whilst also preventing regulatory arbitrage at the expense of European rightholders.

5.3. Like Company v Google and Possible Developments

Any assessment of the legislative trajectory must also take into account the pending preliminary ruling in Like Company v. Google before the CJEU. The case raises fundamental questions about the applicability of the TDM exception to G-AI training and the extraterritorial scope of EU copyright law that are directly relevant to several of the reform proposals examined above. Following the oral hearing of 10 March 2026 — in which the Commission raised questions about the

[180] Lucchi, N., & Hunter, S. (2025), p. 10.
[181] European Parliament resolution of 10 March 2026 (P10_TA(2026)0066), recital Y and para. 12.
[182] EUIPO. (2025), p. 228; Lucchi, N., & Hunter, S. (2025), p. 60.

admissibility of the reference while Like Company framed the case as a structural question about the sustainability of journalism in the age of G-AI — the Advocate General's opinion is expected on 3 September 2026, with a final judgment anticipated by the end of 2026 or early 2027.[183]

The outcome of this case may have a considerable effect on the legislative trajectory outlined above. In the event that the CJEU confirms the broad applicability of Article 4 to G-AI training, the reform proposals examined in this section would operate within a framework that is already largely in place, requiring incremental improvements rather than structural change. Conversely, should the Court adopt a more restrictive interpretation, the pressure for legislative intervention would increase significantly, including the more radical proposals that will be examined in the conclusions.

It is also worth noting in this context that the European Parliament resolution explicitly calls on the Commission to conduct an urgent and comprehensive assessment of whether the current EU copyright acquis adequately addresses the legal uncertainty and competitive effects associated with the use of copyright-protected works for G-AI training, before presupposing the need for legislative revision.[184] This assessment, which is to adopt a holistic approach taking into account the interests of all stakeholders — including researchers, universities, libraries, AI developers, news media, and the creative sector — will be a key determinant of the direction and scope of any future legislative initiative.

[183]European Commission IP Helpdesk, First CJEU hearing on generative AI and copyright: Like Company v Google (24 April 2026), available at https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/cjeu-grand-chamber-rules-music-sampling-and-pastiche-first-cjeu-hearing-generative-ai-and-copyright-2026-04-24_en

[184] European Parliament resolution of 10 March 2026 (P10_TA(2026)0066), para. 15.

CHAPTER III — THE ITALIAN FRAMEWORK: LAW NO. 132/2025 AND ARTICLE 70-SEPTIES OF THE ITALIAN COPYRIGHT ACT

Summary: 1. Introduction to the Italian Framework; 2. TDM Exception within LDA: Articles 70-ter and 70-quater; 3. Legislative Background of Law No. 132/2025; 3.1. Law No. 132/2025: Inspiration and Structure; 3.2. Law No. 132/2025: Legislative Iter; 4. Article 70-septies: Text, Scope and Assessment; 4.1. The Text and Structure of the Provision; 4.2. Critical Assessment of Article 70-septies; 5. The Italian Framework in Context: Unresolved Issues.

1. Introduction to the Italian Framework

The Italian normative framework on copyright in the digital environment has followed European developments. In particular, though not exclusively, through the transposition of the InfoSoc Directive by Legislative Decree of 9 April 2003, No. 68, and of the CDSM Directive by Legislative Decree of 8 November 2021, No. 177.[185] These legislative decrees amended the Italian Copyright Act (Legge 22 aprile 1941, n. 633, hereinafter "LDA"), adapting it to new technological developments. Since the transposed instruments are Directives, national legislation may, within certain limits, depart from the original text — and this has indeed occurred with the legislative decrees referred to above, including, albeit not in a substantial manner, in the transposition of the TDM exceptions (*infra*, §2.)

With respect to G-AI, as discussed in the preceding Chapters, the European legislature opted for a Regulation, directly applicable within the legal orders of the Member States without requiring transposition acts (as a general rule, subject to exceptions). Yet it is precisely in this area that the Italian legislature adopted a legislation — Law of 23 September 2025, No. 132 — to regulate G-AI at the national level, introducing principles and sector-specific provisions that operate alongside the AI Act framework (see below, §2).[186] The Law intervened on the

[185] It should be noted that several other directives have also been transposed into Italian law, including those addressing more specific subject matters such as database protection and software.
[186] Burelli, C. (2026). Commento all'art. 1. In D'Amico, M. et al. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 474-476,

relationship between G-AI, copyright, and the TDM exception introducing the peculiar Article 70-septies LDA. This Chapter, after a brief introduction to the relevant legislative background (Articles 70-ter and 70-quater), aims to analyse the ratio and legislative history of the Law of September 2025, its relationship with the AI Act, and the actual and apparent innovations introduced by the new Article 70-septies LDA, together with the critical issues arising from the new legislative framework.

A methodological observation is warranted at the outset. The legislative instruments analysed in this Chapter — Law No. 132/2025 and the provisions it introduces into the LDA — entered into force on 10 October 2025 and, at the time of writing, have not yet generated judicial decisions or consolidated doctrinal orientations capable of guiding their interpretation. The analysis that follows is therefore necessarily based on the legislative text, its travaux préparatoires, and the limited scholarly commentary available, without the interpretive support that case law and established doctrinal debate would ordinarily provide.

2. TDM Exception within LDA: Articles 70-ter and 70-quater

As anticipated in the preceding section, in transposing Articles 3 and 4 CDSM the Italian legislature modified certain elements of those provisions, generally broadening their scope. Indeed, the EU legislation is a minimum harmonisation directive,[187] and it establishes a minimum standard for the exceptions, but Member States retain latitude to extend their application, within the limits set by Article 25 (i.e. the compatibility with Directives 96/9/EC and InfoSoc Directive) and the three-step test recalled in Article 7 CDSM (*supra*, Chapter I §2.2.2.2).[188] These provisions are the only elements of the pre-existing framework that require specific

noting that the national law risks obscuring the primacy and direct applicability of the AI Act by incorporating, by reference or paraphrase, concepts and principles already established by the Regulation.

[187] Directive CDSM, Article 25: "Member States may adopt or maintain in force broader provisions, compatible with the exceptions and limitations provided for in Directives 96/9/EC and 2001/29/EC, for uses or fields covered by the exceptions or limitations provided for in this Directive."

[188] M. Granieri, 'Il data mining nella disciplina del diritto d'autore e la strategia europea sui dati', AIDA, 2022, §3.

analysis in the Italian context; the considerations already developed at the EU level in Chapters I and II otherwise apply in full.

With respect to Article 70-ter — which contains both the definition of TDM and the TDM exception for scientific research — some departures from the Directive deserve attention. First, the Italian legislature added an express right to communicate to the public the results of research, on the condition that such results are expressed in new original works.[189] This right of public communication is not contemplated in Article 3 CDSM, which is silent on the downstream dissemination of research outputs, and represents a substantive addition by the national legislature.[190] Second, the Italian definition of TDM in Article 70-ter(2) LDA extends the object of the exception beyond the Directive's formulation. While Article 2(2) CDSM refers to 'texts and data in digital format', the Italian legislature specifies “texts, sounds, images, data or metadata in digital format”, thereby encompassing audiovisual and sound materials as well as metadata, which the Directive does not expressly address.[191] Third, It is noteworthy that the Italian definition expressly lists 'models' among the possible outputs of TDM activity. This terminological choice lends support to the argument that the generation of AI models falls within the scope of the TDM exception as originally defined: if the production of models is an expressly contemplated output of text and data mining,

[189] Benatti, F., Le 'nuove' eccezioni e limitazioni al diritto d'autore introdotte dal d.lgs. n. 177/2021, Giurisprudenza commerciale, §4.

[190] Granieri (in M. Granieri, 'Il data mining nella disciplina del diritto d'autore e la strategia europea sui dati', AIDA, 2022, p. 26) argues that the condition limiting the communication to the public to results expressed 'in new original works' is unduly restrictive of the exception: since the extractive activity benefits from the exception at the generative stage, it is difficult to justify why the same protection should not extend to research outputs that cannot be characterised as original works — such as datasets or unstructured data derived directly from the analytical process — but nonetheless fall within the legitimate purposes of the exception. On the other hand, Servanzi (in Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §17, p. 1184) argues that the results of text and data mining will frequently no longer incorporate any protected subject matter; the provision authorising their communication to the public therefore operates as an exception only in the residual cases in which the research output still contains protected subject matter, since it is only in such cases that an authorisation would otherwise be required — absent a protected work, no exception would be necessary in the first place. Always following the reasoning of the author, this reading would raise questions in relation to the EU compatibility of the provision, as in view of the principle, Member States may not introduce exceptions to copyright beyond those permitted by the Directives.

[191] Benatti, F., Le 'nuove' eccezioni e limitazioni al diritto d'autore introdotte dal d.lgs. n. 177/2021, Giurisprudenza commerciale, §4.2.

it becomes difficult to maintain that the training of AI models — as a species of that activity — was not already covered by the exception prior to the introduction of Article 70-septies. This raises a question of some interpretive significance: if the generation of models was already an expressly contemplated output of text and data mining under the Italian definition, it may be asked whether Article 70-septies, in extending the TDM framework to G-AI training, introduces a genuinely new element into the legal framework or whether it operates principally by way of confirmation of what the existing provisions may already have implied. Finally, the Italian definition introduces a quantitative threshold, limiting the scope of the exception to analyses conducted on 'large quantities' of material, a criterion absent from the Directive and undefined in the national text.[192]

With respect to Article 70-quater — to which the observations concerning the definition of TDM in Article 70-ter equally apply — the principal divergence from the CDSM framework concerns the opt-out mechanism. Article 70-quater(1) LDA conditions the exception on the absence of an express reservation by rights holders, providing that reproductions and extractions are permitted "when the use of the works and other materials has not been expressly reserved by the holders of copyright and related rights and by database right holders". This formulation diverges from Article 4(3) CDSM, which requires that rights holders must have reserved their rights "in an appropriate manner". The concept of appropriate reservation is clarified by Recital 18 CDSM, which distinguishes between content made available online and content made available through other means: for online content, rights holders may validly reserve their rights only through machine-readable means, including metadata and the terms and conditions of a website or service; for content made available in other ways, reservation through contractual clauses or unilateral declarations is also permissible. In any event, this divergence does not improve upon, nor in any way modify, the difficulties identified in Chapter II §2.2 with respect to the EU opt-out framework. It should be noted,

[192] Ibidem.

however, that the omission of an express "appropriate manner" qualification in Article 70-quater(1) LDA is not generally regarded by Italian scholarship as producing a substantive divergence from the Directive. The prevailing view holds that the requirement of adequacy — and, for content made available online, of machine-readability — must be read into the national provision by way of interpretation, in light of Recital 18 CDSM and of the function that Article 4 assigns to the reservation, namely that of ensuring an application that is simple enough to incentivise mining activity: it follows that a reservation which cannot be interpreted by automated systems is, on this reading, ineffective, with the correlative consequence that the extraction remains permitted.[193] The national omission, therefore, does not in itself enlarge or diminish the scope of the opt-out as compared with the Directive. What it does not resolve, however, are the practical difficulties of fragmentation, machine-readability standards, and overlapping declarations examined in Chapter II §2.2, which persist in the Italian context precisely because the interpretive reconstruction of the adequacy requirement does not, of itself, supply the technical and standard-setting answers that the operability of the mechanism requires.

As regards both provisions, a further observation is warranted: neither Article 70-ter nor Article 70-quater LDA contains an express reference to the three-step test. This is probably a symptom of the fact that when the exception was transposed the legislator did not perceive a possible friction between the TDM and the three-step test. However, as analysed in Chapter I §2.2.2.2, the three-step test applies to these exceptions by virtue of Article 7(2) CDSM and Article 5(5) InfoSoc, irrespective of whether it has been expressly reproduced in national law, and national courts are bound to apply it accordingly. Moreover, in relation to G-AI training, the Italian legislature felt the necessity to require conformity with the Berne Convention, which may relate also to three-step compliance (*infra*, §4.1.)

[193] In this sense Servanzi, R., Le estrazioni di testo e di dati, Le Nuove Leggi Civili Commentate, 2022, §10, pp. 1174–1175, citing Ottolia, A., L'opt out commons nella nuova disciplina del data mining.

3. Legislative Background of Law No. 132/2025

3.1. Law No. 132/2025: Inspiration and Structure

On 25 September 2025, Italy published Law No. 132 of 23 September 2025, entitled 'Provisions and Delegations to the Government on Artificial Intelligence', which entered into force on 10 October 2025.[194] With this enactment, Italy became the first Member State of the European Union to adopt a comprehensive national legislative framework for artificial intelligence, preceding the full applicability of Regulation (EU) 2024/1689 (the AI Act), which does not apply in full until 2 August 2026.[195]

The timing of the Italian intervention reflects a deliberate political choice. The law was approved while several implementing acts of the AI Act — including delegated acts and codes of practice — had not yet been adopted, and while the full operational architecture of the EU regulatory framework was still being constructed.[196] The decision to legislate in advance of the complete entry into force of the AI Act responds to the urgency of regulating a sector of considerable economic and social relevance.[197] Federazione Italiana Editori Giornali (FIEG) also noted that Italy had repeatedly expressed, including at the G7 level, a position in favour of strong copyright protection in the context of the AI Act negotiations.[198]

The coexistence of Law No. 132/2025 with the AI Act within the same legal order raises questions of regulatory coordination. The AI Act is a Regulation, and as such it is directly applicable in all Member States without the need for implementing acts. In order to address the issue of potential normative conflict, the

[194] Legge 23 settembre 2025, n. 132, Gazzetta Ufficiale n. 223 del 25 settembre 2025.
[195] AI Act, Art. 113.
[196] Burelli, C. (2026). Commento all'art. 1. In D'Amico, M. et al. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, p. 471.
[197] Analisi di impatto della regolamentazione (AIR) relativa al disegno di legge A.S. 1146, 2024, p. 6-7, available at https://www.senato.it/service/PDF/PDFServer/BGT/01419746.pdf.
[198] FIEG/ENPA, Document on DDL 1146 (April 2024), p. 1. The original Italian reads: 'il tema del copyright è risultato al centro del dibattito e l'Italia ha più volte evidenziato una posizione favorevole a una forte tutela dei diritti d'autore, posizione ribadita anche a livello di G7.'

Italian legislature established two coordination mechanisms. Article 1(2) L. 132/2025 establishes the requirement for its provisions to be interpreted and applied in conformity with Regulation (EU) 2024/1689. Furthermore, Article 3(5) provides that the law does not create new obligations beyond those which have been established by the AI Act. These clauses are indicative of a national law that is positioned as complementary to, rather than in conflict with, the EU Regulation. Nevertheless, legal scholarship has identified a structural risk in this legislative technique. By incorporating, by reference or by paraphrase, concepts and principles already established by EU Regulations, national law risks obscuring the primacy and direct applicability of those instruments. This is a concern grounded in established CJEU case law requiring that Member States do not reproduce or paraphrase Regulation provisions in ways that dissimulate their nature as Union law.[199]

The law is structured in six chapters: Chapter I sets out principles and purposes; Chapter II contains sector-specific provisions covering healthcare, labour, public administration and judicial activity; Chapter III addresses national strategy, national authorities and promotional actions; Chapter IV introduces provisions for the protection of users and copyright; Chapter V introduces criminal law amendments; and Chapter VI contains financial and final provisions. For the purpose of this thesis Chapter IV is of particular relevance, as it focuses on the intersection between G-AI training, copyright and TDM exception.

As discussed in Chapter II, the AI Act does not directly intervene on copyright, but confines itself to a cross-reference, through Article 53, to the CDSM Directive. Article 70-septies LDA, by contrast, does not address aspects subsequent to the training process itself — such as transparency obligations or compliance policies

[199]Burelli, C. (2026). Commento all'art. 1. In D'Amico, M. et al. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 474-476, citing CJEU, Commissione c. Italia, Case 39/72, EU:C:1973:13, para. 17; Commissione c. Danimarca, C-541/16, EU:C:2018:251, para. 28. The author notes that the overlap of purposes creates a risk of 'oscuramento' of the primary normative source — the EU Regulation — and that the national law risks dissimulating its nature as an act of Union law.

— but intervenes directly on the permissibility of conducting training through the reproduction and extraction of works and other materials, subject to conformity with the national TDM provisions and without prejudice to the Berne Convention. The reading of this provision raises questions as to its actual impact — whether substantive or interpretive — which will be examined in the following sections, first by tracing its legislative history and subsequently by analysing its effects.

3.2. Law No. 132/2025: Legislative Iter

The legislative history of what became Article 70-septies spans from the Council of Ministers' approval of DDL 1146 on 23 April 2024 to the publication of the final law in September 2025. The DDL was presented to the Senate on 20 May 2024, approved by the Senate on 20 March 2025, passed by the Chamber with further amendments on 25 June 2025, and definitively approved by the Senate on 17 September 2025.[200] The parliamentary iter was characterised by significant modifications to the copyright provisions, driven by a combination of external observations — from the European Commission and from industry stakeholders — and internal parliamentary debate.

In the original government draft, what would subsequently become Article 70-septies was included in Article 24(1)(b) DDL as a bare cross-reference: reproductions and extractions of works and materials through AI models and systems, including generative AI, are permitted when performed in conformity with Articles 70-ter and 70-quater LDA.[201] The formulation of this particular Article did not contain independent substantive requirements, nor did it make reference to the Berne Convention, and no explicit lawful access condition was stipulated. The provision functioned as a pure normative cross-reference,

[200] Camera dei deputati, Dossier ID0010c, giugno 2025; Senato della Repubblica, Relazione BGT/1461539, giugno 2025.

[201] DDL A.S. 1146, testo d'iniziativa governativa, Art. 24(1)(b). The original Italian text reads: 'La riproduzione e l'estrazione di opere o altri materiali attraverso modelli e sistemi di intelligenza artificiale anche generativa sono consentite in conformità alle disposizioni di cui agli articoli 70-ter e 70-quater.'

extending the existing TDM framework to G-AI training without the addition of new content.

This minimal character attracted immediate criticism. FIEG, who was also acting on behalf of the European Newspapers Publishers' Association (ENPA), submitted observations in April 2024. In these observations, FIEG explicitly stated that "The mere cross-reference to the Text and Data Mining articles appears almost superfluous and, in any event, not decisive."[202] FIEG/ENPA identified a series of more substantive interventions that had been proposed in earlier drafts. These included the simplification of opt-out procedures, incentives for collective licensing models, registration obligations for works used in AI training, and the recognition of equitable remuneration for rights holders. However, the text under examination did not include any of these interventions. The observations further propose two specific modifications: the scope of the provision should be limited to 'models' in accordance with Recital 107 AI Act; and it should be clarified that the exception applies only to activities constituting TDM in the strict sense, excluding uses that go beyond pattern analysis towards reproduction of copyrightable expression. It is evident that neither proposal was incorporated into the final text.

On 5 November 2024, the European Commission transmitted a detailed opinion pursuant to Article 6(2) of Directive (EU) 2015/1535 on technical regulations, raising concerns about the compatibility of several provisions of DDL 1146 with EU law.[203] The most consequential outcome for the copyright chapter was the suppression of what was originally Article 23 DDL — a provision requiring visible identification of AI-generated content with an 'IA' acronym — following the Commission's finding that it overlapped with and went beyond the disclosure

[202] FIEG/ENPA, Document on DDL 1146 (April 2024), p. 5. The original Italian reads: 'il mero richiamo agli articoli sul Text and Data Mining appare quasi superfluo e comunque non decisivo.'
[203] European Commission, Detailed Opinion pursuant to Article 6(2) of Directive (EU) 2015/1535, C(2024) 7814, 5 November 2024, as reported in Senato della Repubblica, N. 1146-A, parere della 4a Commissione permanente (Politiche dell'Unione Europea), 27 novembre 2024.

obligations of Article 50 AI Act. On the copyright provisions specifically, the Commission did not raise objections to the cross-referential approach per se, but its broader message — that national interventions should avoid creating overlap or inconsistency with EU instruments — informed the subsequent national parliamentary debate and the observations of the 4a Commission (EU Policies), which specifically invited the committees of merit to consider whether definitions in the national law should simply refer to those in the AI Act rather than attempting autonomous formulations.[204] The 2a Commission (Justice) additionally raised the question of whether the copyright provisions, given their systemic impact on national copyright law, should have been examined independently rather than embedded in a broader AI governance statute — a suggestion that was not followed.[205]

4. Article 70-septies: Text, Scope and Assessment

4.1. The Text and Structure of the Provision

The final text of Article 70-septies LDA, as introduced by Article 25(1)(b) L. 132/2025, reads as follows: "Without prejudice to the provisions of the Berne Convention for the Protection of Literary and Artistic Works, ratified and made enforceable pursuant to Law No. 399 of 20 June 1978, reproductions and extractions from works or other materials contained in networks or databases to which lawful access has been obtained, for the purpose of text and data mining through artificial intelligence models and systems, including generative AI, are permitted in conformity with the provisions of Articles 70-ter and 70-quater."[206]

[204] Senato della Repubblica, N. 1146-A, parere della 4a Commissione permanente (Politiche dell'Unione Europea), 27 novembre 2024.
[205] Senato della Repubblica, N. 1146-A, parere della 2a Commissione permanente (Giustizia), 20 novembre 2024, pp. 8-9.
[206] Legge 22 aprile 1941, n. 633, Art. 70-septies, as introduced by Art. 25(1)(b) L. 132/2025. The original Italian text reads: 'Fermo restando quanto previsto dalla Convenzione di Berna per la protezione delle opere letterarie ed artistiche, ratificata e resa esecutiva ai sensi della legge 20 giugno 1978, n. 399, le riproduzioni e le estrazioni da opere o da altri materiali contenuti in rete o in banche di dati a cui si ha legittimamente accesso, ai fini dell'estrazione di testo e di dati attraverso modelli e sistemi di intelligenza artificiale, anche generativa, sono consentite in conformità alle disposizioni di cui agli articoli 70-ter e 70-quater.'

Article 70-septies is composed of a single paragraph and can be analysed by reference to its constitutive elements.

The protected activity encompasses the reproduction and extraction of works or other materials. The focus of this directive is on text and data mining, as delineated in Article 70-ter(2) LDA, given the absence of an independent definition. The directive stipulates that this mining should be conducted through the utilisation of AI models and systems, encompassing generative AI. This formulation, as elaborated in the subsequent discussion, gives rise to specific inquiries concerning its extent.

With regard to the subjective scope, the provision contains no limitation as to who may rely on it. In contrast to Article 70-ter, which is exclusively available to research organisations and cultural heritage institutions, Article 70-septies is accessible to any individual, whether they are a natural person, a legal entity, a public body or a private entity, or a commercial entity or a non-commercial entity. However, it should be noted that the applicable conditions are not uniform. The provision requires conformity with Articles 70-ter and 70-quater, and thus the relevant conditions will differ depending on which of these two exceptions the activity falls under. As explained above, Article 70-ter applies only to scientific research by qualifying entities, while Article 70-quater is more broadly applicable, but subject to the opt-out mechanism.

The conditions to which the permission is subject are three. First, the activity must be conducted without prejudice to the Berne Convention — a savings clause whose significance is examined below. Second, the reproductions and extractions must be made from materials to which lawful access has been obtained — a condition that raises the question of whether it adds independent content, given that the cross-reference to Articles 70-ter and 70-quater already makes lawful access a condition of those provisions. Third, and most fundamentally, the activity must be conducted in conformity with Articles 70-ter and 70-quater — a cross-reference that defines

the operative content of the permission and gives rise to the central question addressed in the following section: whether this normative technique produces any genuine effect beyond what the existing TDM framework already provided.

The provision presents several elements worthy of analysis. The first is the reference to the Berne Convention, which — as noted above — is absent from both Articles 70-ter and 70-quater. The travaux préparatoires offer no explanation for this addition, which was not present in the original DDL text. The Berne Convention represents one of the principal international instruments for copyright protection and rests on several foundational principles, including the principle that the author who has created a work is its owner, the principle of automatic protection under which rights arise at the moment of creation without any administrative formality, the principle of national treatment, and the principle of independence of protection. The savings clause "without prejudice to the provisions of the Berne Convention" thus invokes the full corpus of the Convention, preserving the applicability of all these principles to the activity of G-AI training through AI models and systems.[207]

Among the principles enshrined in the Convention, one of particular relevance in the context of the present thesis — in light of the doctrinal debate examined in Chapter I §2.2.2.2. — is the three-step test of Article 9(2) Berne. It may be observed that, since Article 70-septies is not a transposition of the CDSM Directive — which already integrates the three-step test through Article 7 — the Italian legislature may have felt it appropriate to make express reference to the Convention in order to ensure that this standard applies equally to the new provision. It is nonetheless worth noting that elsewhere in the LDA, when the legislature has wished to invoke the Berne Convention with specific reference to the three-step test, it has done so in explicit terms: Articles 64-quater(4) and 64-

[207] Panigada, N. (2026). Commento all'art. 25. In D'Amico, M., Balaguer Callejón, F., Aguilar Calahorro, A., Gambatesa, P., Di Giovanni, S. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 669-670.

sexies(4) both contain language that reproduces substantially the three conditions of the test, providing that those provisions 'shall not be interpreted in a way that allows their application to cause unreasonable prejudice to the legitimate interests of the rightholder or to conflict with the normal exploitation' of the protected subject matter. The general savings clause adopted in Article 70-septies departs from that more targeted drafting technique.

The savings clause in favour of the Berne Convention may, moreover, be read as engaging a dimension of the Convention that is analytically distinct from the three-step test, namely the prohibition of constitutive formalities under Article 5(2). As examined in Chapter II §2.4, Italian scholarship has questioned whether the reservation mechanism on which Articles 70-quater and 4(3) CDSM are built is compatible with that prohibition, at least as regards non-EU rightholders, in respect of whom the obligations deriving from EU law are not opposable and the directly applicable Convention standard prevails. Read against this background, however, the express invocation of the Convention does not resolve the difficulty. Article 70-septies operates exclusively by cross-reference to Articles 70-ter and 70-quater, and thereby reintroduces, for the field of G-AI training, the very reservation mechanism of Article 70-quater whose compatibility with the prohibition of constitutive formalities is in question. The reference to Berne and the cross-reference to Article 70-quater thus pull in opposite directions, and the former does not neutralise the latter: notwithstanding the invocation of a Convention that prohibits constitutive formalities, the opt-out returns through the cross-reference, and with it the very question of compatibility that the reference to the Convention might have been expected to resolve.

The second element is the lawful access condition, equally absent from the original DDL text. It may be questioned whether this condition adds any independent content to the provision. Since conformity with Articles 70-ter and 70-quater is required, and both of those provisions already make lawful access a condition of their application, the explicit repetition of the requirement in Article 70-septies

raises the interpretive question of whether it was intended to produce any effect beyond that already implied by the cross-reference.

The third element is the extension of the provision's scope to both AI models and AI systems. In the terminology of Law No. 132/2025, these are distinct categories. Article 2(1)(c) L. 132/2025 defines "AI models" by reference to Article 3(63) AI Act — the definition of "general-purpose AI model" — thereby equating the national category of AI model with GPAI models as defined by EU law. "AI systems", by contrast, are defined by Article 3(1) AI Act as automated systems with varying degrees of autonomy capable of generating outputs that influence their environment. The concept of "AI models" in national law is thus functionally limited to general-purpose AI models, which constitutes a narrower and distinct category from "AI systems".[208] By extending Article 70-septies to cover both, the legislature ensures that the provision applies not only to the training of large foundational models, but also to AI systems — which may include downstream applications built on top of those models. Whether this extension represents a deliberate policy choice or an inadvertent consequence of the definitional structure of the national law is not entirely clear from the travaux preparatoires. What is clear is that it expands the potential scope of the provision beyond what would have been covered by a reference to "models" alone, and goes beyond Recital 107 AI Act, which refers specifically to "models" rather than "systems" in the context of copyright compliance obligations.[209]

The fourth and most significant element is the cross-reference to Articles 70-ter and 70-quater, which define the operative content of the permission. Article 70-septies stipulates that reproductions and extractions for G-AI training are permitted in accordance with the provisions outlined therein. This normative technique serves to extend the scope of an existing legal regime to a novel category of

[208] Burelli, C. (2026). Commento all'art. 1. In D'Amico, M. et al. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 480-481
[209] Regulation (EU) 2024/1689, Recital 107.

activity. The following section will examine whether this extension produces normative content that was not already implicit in the existing framework, or whether it operates principally as a legislative clarification of the applicability of that framework to G-AI training.

4.2. Critical Assessment of Article 70-septies

The structural analysis of Article 70-septies in §4.1 raises a question of interpretive significance: whether the provision produces any normative content that was not already implicit in the existing framework, or whether it operates principally as a confirmation of the applicability of that framework to G-AI training. The question is not without practical consequences. As noted above, each of the three conditions imposed by the article — the Berne savings clause, the lawful access requirement, and the cross-reference to Articles 70-ter and 70-quater — corresponds to obligations or conditions that were already operative in the Italian legal order. The Berne savings clause restates obligations binding Italy through its ratification of the Convention pursuant to Law No. 399/1978. The lawful access condition replicates a requirement already established by both Articles 70-ter and 70-quater. The cross-reference, finally, extends the existing TDM regime to a new category of activity without modifying its content.

These observations give rise to the question of whether Article 70-septies should be interpreted as constitutive — that is, as establishing for the first time that G-AI training falls within the scope of the TDM exceptions — or as confirmatory of what was already implicit in the existing framework. The two readings yield divergent outcomes. In the context of the constitutive reading, the introduction of a specific provision addressing AI training could be interpreted as an implicit acknowledgement that the pre-existing framework was not already comprehensive with respect to this activity. This would consequently give rise to questions concerning the legal position under Italian law prior to October 2025. In the event of confirmation, the provision would operate as a legislative clarification within a

context of legal uncertainty. In this way, the Italian legislature's interpretive position would be placed on record without any alteration to the substantive rules.

The legislative history examined in §3.2 offers some, albeit limited, guidance on this question. FIEG/ENPA, commenting on the original DDL text, observed that the bare cross-reference to the TDM articles appeared "almost superfluous and in any event not decisive", and called for more substantive interventions that were not ultimately incorporated in the final text.[210] This critical assessment, advanced by the principal stakeholder representing rights holders in the legislative process, suggests that the cross-referential approach was not regarded as introducing substantively new content even from the perspective of those most attentive to its normative impact.

Nevertheless, the provision accomplishes a result that the pre-existing framework alone could not guarantee: by explicitly and unambiguously stating that reproductions and extractions for G-AI training are permitted in conformity with Articles 70-ter and 70-quater, the Italian legislature has placed its interpretive position on record at the national level. This has the effect of limiting, within the Italian legal order, the scope for arguing that the TDM framework was not intended to cover G-AI training — an argument whose force at the EU level, where the question of legislative intent behind Articles 3 and 4 CDSM remains a matter for the CJEU, is unaffected by the Italian provision.

5. The Italian Framework in Context: Unresolved Issues

The Italian normative framework — combining the LDA as amended with the directly applicable AI Act — does not resolve the structural problems identified in the preceding chapters. Whether G-AI training is conducted under Articles 70-ter or 70-quater as extended by Article 70-septies, or directly under those provisions, the applicable conditions and their operative difficulties remain unchanged. The challenges surrounding the opt-out mechanism, analysed in Chapter II §2.2, apply

[210] FIEG/ENPA, Document on DDL 1146 (April 2024), p. 5

in full in the Italian context: Article 70-septies incorporates Article 70-quater by reference without addressing any of the practical difficulties of fragmentation, machine readability, and overlapping declarations that make the mechanism inadequate at the scale of G-AI training. Similarly, the interpretive questions surrounding lawful access, examined in Chapter II §2.3, are reproduced without modification in the national context. The Italian legislature, having chosen to introduce a specific provision on AI training, provided no national guidance on any of these questions. The conditions applicable under Article 70-septies are those already established by the EU framework — and that framework, as demonstrated in Chapters I and II, does not yet provide adequate answers to the practical questions raised by G-AI training at scale.

CONCLUSIONS

The central question animating this thesis — whether generative AI training can be accommodated within the existing EU copyright framework, and under what conditions — does not admit of a simple answer. The analysis conducted in the preceding chapters reveals a legal landscape that is, at once, more structured than the public debate sometimes suggests and more uncertain than the current regulatory framework can comfortably sustain.

It is an established fact that G-AI training involves acts of reproduction for the purposes of Article 2 of the InfoSoc Directive at the stage of content ingestion. The

CJEU has confirmed that the copying, storing, and pre-processing of training data are paradigmatic cases of reproduction, as evidenced by the broad interpretation of Article 2 InfoSoc developed by the Court. The extraction right under Directive 96/9/EC is also engaged in cases where training datasets are assembled from protected databases. The question that remains the subject of genuine contestation is whether a trained AI model constitutes, in itself, a reproduction of the copyright-protected works used during the training process. The discord that has surfaced in academic discourse and legal precedent on this issue is not merely a technical dispute concerning the encoding of information by model parameters; rather, it is a methodological disagreement regarding the prioritisation of legal inquiries. One line of argument focuses on the model's demonstrable capacity to regenerate protected expression as evidence of legally relevant retention, a position that finds some support in recent German case law. A different approach insists that the analysis must begin with whether a copyright-protected work is identifiable with sufficient precision and objectivity within the model itself — a condition that, on the CJEU's own case law on the notion of a work, is not satisfied in relation to AI model parameters. This thesis has argued that the latter reading is both more doctrinally coherent and more consistent with the systematic connection between Article 4 CDSM and Article 53 AI Act, which presupposes that the trained model is not itself a reproduction of the training data. Technical developments aimed at mitigating memorisation reinforce this conclusion by progressively reducing the factual basis for the opposite argument. The question, however, has not been definitively resolved by the CJEU, and the pending preliminary ruling in Like Company v. Google offers the Court an opportunity to provide authoritative guidance on precisely this point.

With regard to the TDM exception, the analysis has demonstrated that G-AI training incorporates, at least in part, techniques that qualify as text and data mining within the meaning of Article 2(2) CDSM. The automated analysis of large quantities of digital content for the purpose of generating information – including,

in the terminology of the Directive and its Italian transposition, models – falls within the broad definition of TDM. It is evident that the AI Act's reference to TDM in the context of GPAI training in Recital 105 AI Act, and its explicit reliance on the opt-out mechanism of Article 4(3) CDSM as the primary instrument of copyright compliance for GPAI providers, constitutes strong systematic evidence that the EU legislature regarded GPAI training as falling within the scope of the TDM exception framework. The teleological argument – that Articles 3 and 4 of the CDSM were not intended to cover AI training – is difficult to sustain in light of this legislative context. The more challenging question is not whether the TDM exception applies, but whether it applies in an adequate manner. The three-step test, as interpreted by the CJEU and applied through Article 7 CDSM, gives rise to questions regarding the compatibility of large-scale commercial G-AI training with the exception framework that scholarship has yet to resolve. The requirement of a sufficiently narrow special case, the prohibition of conflict with the normal exploitation of the work, and the requirement that the legitimate interests of rights holders not be unreasonably prejudiced each give rise to genuine interpretive difficulties in the G-AI context. While the opt-out mechanism of Article 4(3) CDSM provides a balancing mechanism that is specifically designed to address these concerns, it does so only to the extent that it is effectively operational. The analysis in Chapter II has demonstrated that it might not be the case: the fragmentation of opt-out methods, the absence of a universally recognised machine-readable standard, the localisation problem, and the overlapping rights structure collectively render the mechanism inadequate as a practical guarantee of rights holders' interests at the scale of G-AI training. Beyond these operational shortcomings, the analysis has flagged a question of a different order — whether the reservation mechanism is itself compatible with the prohibition of constitutive formalities under Article 5(2) of the Berne Convention — which would bear not on the practical effectiveness of the opt-out but on its legitimacy as such.

The AI Act does not harmonise copyright and does not create or modify substantive copyright exceptions. Its contribution to the field is procedural and regulatory in character: it imposes on GPAI providers obligations to adopt and maintain a copyright compliance policy, to identify and respect opt-outs, and to publish a summary of training data. The Code of Practice for General-Purpose AI Models operationalises these obligations through a more detailed set of measures. These are not without significance: by requiring GPAI providers to engage systematically with the opt-out mechanism, the AI Act creates institutional pressure toward the development of more standardised and interoperable reservation protocols. At the same time, the Code of Practice presents structural limitations that have been examined in Chapter II — its application is confined to web-crawled data, it addresses only GPAI providers, and it relies on self-reporting mechanisms whose accuracy cannot be independently verified. The practical consequence is that the burden of substantive copyright enforcement in the G-AI training context falls primarily on rights holders, who must independently establish infringement through private litigation — typically without access to the information necessary to do so. The information asymmetry between rights holders and GPAI providers is mitigated but not resolved by the AI Act framework.

With respect to the Italian legislative framework, Article 70-septies LDA, introduced by Law No. 132/2025, extends the existing TDM framework to G-AI training through a cross-referential technique, without introducing autonomous conditions beyond those already established by Articles 70-ter and 70-quater. As examined in §4.2, questions arise as to the actual normative contribution of the provision: each of its three operative elements — the Berne savings clause, the lawful access condition, and the cross-reference to the existing TDM exceptions — corresponds to obligations or conditions already operative in the Italian legal order. The provision does, however, place the Italian legislature's interpretive position on record, limiting the scope for arguing within the Italian legal order that the TDM framework was not intended to cover G-AI training. More broadly, the

Italian framework illustrates a recurring challenge for national legislators operating in a field largely shaped by EU law: the desire to provide a visible national response to a pressing regulatory question may result in provisions that reproduce at the national level the same open questions that characterise the EU framework.

The analysis conducted in this thesis leads to three principal conclusions. First, G-AI training can, at least in part, be accommodated within the existing EU copyright framework through the TDM exception, provided that the conditions of lawful access and opt-out compliance are satisfied. Whether large-scale commercial G-AI training satisfies those conditions in practice depends on empirical and normative assessments that cannot be made in the abstract and that remain contested in scholarship and case law. Second, the existing framework is structurally inadequate to ensure effective rights holder protection in the G-AI training context: the opt-out mechanism, the transparency obligations, and the enforcement architecture collectively fail to provide rights holders with the tools necessary to exercise their rights effectively against GPAI providers operating at scale. Third, the Italian legislative intervention extends the existing TDM framework to G-AI training through a cross-referential technique, without introducing autonomous conditions, and leaves open at the national level the same interpretive questions that characterise the EU framework.

The trajectory of EU copyright law in this field is likely to be shaped, in the near term, by three developments. The CJEU's ruling in Like Company v. Google will provide — for the first time — authoritative judicial interpretation of the TDM exception's application to G-AI training, with consequences that may either stabilise the existing framework or create pressure for legislative reform. The Commission's assessment, called for by the European Parliament's resolution of 10 March 2026, will determine whether and in what direction legislative revision is necessary. And the ongoing development of standardised machine-readable opt-out protocols — through the Code of Practice process and the proposed EUIPO

registry mechanism — will determine whether the opt-out mechanism can be made operationally effective without legislative change. What this thesis has ultimately sought to demonstrate is that the question of G-AI training and copyright is neither as simple as those who argue for a blanket application of the TDM exception suggest, nor as intractable as those who argue for a complete prohibition would have it. It is a question that requires careful legal analysis, attentiveness to the rapidly evolving technical context, and a willingness to engage with the difficult trade-offs between innovation and rights protection that have always been at the heart of EU copyright law. The existing framework provides a starting point — but not a sufficient answer.

LEGISLATION

I. INTERNATIONAL LEGISLATION

Berne Convention for the *Protection of Literary and Artistic Works* (1886, as revised)

Agreement on *Trade-Related Aspects of Intellectual Property Rights* (TRIPS), Annex 1C to the Marrakesh Agreement, 15 April 1994

Marrakesh Agreement Establishing the World Trade Organization, 15 April 1994, entered into force 1 January 1995

WIPO Copyright Treaty (WCT), adopted in Geneva on 20 December 1996; approved on behalf of the European Community by Council Decision 2000/278/EC of 16 March 2000, OJ L 89, 11.4.2000, pp. 6–7

Agreed Statements concerning the WIPO Copyright Treaty, adopted by the Diplomatic Conference on 20 December 1996

OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019)

WTO Panel Report, United States — Section 110(5) of the US Copyright Act, WT/DS160/R, 15 June 2000

II. EU LEGISLATION

Council Directive 91/250/EEC of 14 May 1991 *on the legal protection of computer programs*, OJ L 122, 17.5.1991, pp. 42–46

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 *on the legal protection of databases*, OJ L 77, 27.3.1996, pp. 20–28

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* (InfoSoc Directive), OJ L 167, 22.6.2001, pp. 10–19

Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the *enforcement of intellectual property rights* (Enforcement Directive), OJ L 157, 30.4.2004, pp. 45–86

Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 *on the legal protection of computer programs* (Software Directive), OJ L 111, 5.5.2009, pp. 16–22

Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 *laying down a procedure for the provision of information in*

the field of technical regulations and of rules on Information Society services (codification) (Text with EEA relevance), OJ L 241, 17.9.2015, pp. 1–15.

Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 *on the protection of undisclosed know-how and business information* (Trade Secrets Directive), OJ L 157, 15.6.2016

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 (CDSM Directive), OJ L 130, 17.5.2019, pp. 92–125

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, pp. 56–83

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 *laying down harmonised rules on artificial intelligence* (AI Act), OJ L, 12.7.2024

Council Decision 94/800/EC of 22 December 1994 concerning *the conclusion on behalf of the European Community of the agreements reached in the Uruguay Round multilateral negotiations*, OJ L 336, 23.12.1994, pp. 1–2

III. ITALIAN LEGISLATION

Legge 22 aprile 1941, n. 633, *Protezione del diritto d'autore e di altri diritti connessi al suo esercizio (LDA)*, in Gazzetta Ufficiale n. 166, 16 luglio 1941, come modificata, da ultimo, dal decreto legislativo 8 novembre 2021, n. 177 e dalla legge 23 settembre 2025, n. 132.

Decreto legislativo 9 aprile 2003, n. 68, *Attuazione della direttiva 2001/29/CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione*, in Gazzetta Ufficiale n. 87, 14 aprile 2003

Decreto legislativo 8 novembre 2021, n. 177, Attuazione della direttiva (UE) 2019/790 del Parlamento europeo e del Consiglio, del 17 aprile 2019, sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE, in Gazzetta Ufficiale n. 283, 27 novembre 2021.Legge 20 giugno 1978, n. 399 (ratifica della Convenzione di Berna)

Legge 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale, Gazzetta Ufficiale n. 223 del 25 settembre 2025

Disegno di legge A.S. 1146 (testo d'iniziativa governativa, approvato dal Consiglio dei Ministri il 23 aprile 2024)

IV. UK LEGISLATION

Copyright, Designs and Patents Act 1988 (UK), ss. 22, 23, 27(3)

V. INSTITUTIONAL DOCUMENTS AND WORKING PAPERS

Commission of the European Communities, *Green Paper on Copyright and the Challenge of Technology — Copyright Issues Requiring Immediate Action*, COM(88) 172 final, 1988

European Commission, *Green Paper on Copyright and Related Rights in the Information Society*, COM(95) 382 final, 27 July 1995

European Commission, *Reflection Document of DG INFSO and DG MARKT on Creative Content in a European Digital Single Market: Challenges for the Future*, 22 October 2009

European Commission, Communication: *A Digital Single Market Strategy for Europe,* COM(2015) 192 final

European Commission, Communication: T*owards a Modern, More European Copyright Framework*, COM(2015) 626 final

European Commission, *Proposal for a Directive on Copyright in the Digital Single Market*, COM/2016/593 final

European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final; with accompanying Impact Assessment, SWD(2021) 84 final

European Commission, Detailed Opinion pursuant to Article 6(2) of Directive (EU) 2015/1535, C(2024) 7814, 5 November 2024

European Parliament, Resolution of 20 October 2020 *on intellectual property rights for the development of artificial intelligence technologies*, 2020/2015(INI), OJ C 404, 6.10.2021, p. 129

European Parliament, Amendments adopted on 14 June 2023 on *the proposal for a regulation laying down harmonised rules on artificial intelligence*, P9_TA(2023)0236

European Parliament, Resolution of 10 March 2026 *on copyright and generative artificial intelligence*, P10_TA(2026)0066

Code of Practice for General-Purpose AI Models, Copyright Chapter (2025), entered into force 2 August 2025

European Commission IP Helpdesk, *First CJEU hearing on generative AI and copyright: Like Company v Google*, 24 April 2026, available at https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/cjeu-grand-chamber-rules-music-sampling-and-pastiche-first-cjeu-hearing-generative-ai-and-copyright-2026-04-24_en

European Commission: Directorate-General for Communications Networks, Content and Technology; Technopolis Group; UCLouvain; IMC University of Applied Sciences Krems; Crowell & Moring; Philippe Rixhon Associates, *Study on copyright and new technologies – Copyright data management and*

artificial intelligence, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2759/570559

Analisi di impatto della regolamentazione (AIR) relativa al disegno di legge A.S. 1146, 2024, available at https://www.senato.it/service/PDF/PDFServer/BGT/01419746.pdf

FIEG/ENPA, Observations on DDL 1146, April 2024

Senato della Repubblica, N. 1146-A, Parere della 4a Commissione permanente (Politiche dell'Unione Europea), 27 novembre 2024

Senato della Repubblica, N. 1146-A, Parere della 2a Commissione permanente (Giustizia), 20 novembre 2024

Camera dei deputati, Dossier ID0010c, giugno 2025

Senato della Repubblica, Relazione BGT/1461539, giugno 2025

CASE LAW

I. EU COURT OF JUSTICE

CJEU, *EMI Electrola GmbH v Patricia Im- und Export Verwaltungsgesellschaft mbH*, C-341/87, EU:C:1989:30

CJEU, *Infopaq International A/S v Danske Dagblades Forening* (Infopaq I), C-5/08, EU:C:2009:465

CJEU, *Infopaq International A/S v Danske Dagblades Forening* (Infopaq II), C-302/10, EU:C:2012:16

CJEU, *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg*, C-304/07, EU:C:2008:552

CJEU, *Bezpečnostní softwarová asociace — Svaz softwarové ochrany v Ministerstvo kultury*, C-393/09, EU:C:2010:816

CJEU, *Football Association Premier League Ltd and Others v QC Leisure and Others*, C-403/08 and C-429/08, EU:C:2011:631

CJEU, *Ulmer*, C-117/13, EU:C:2014:2196

CJEU, *ACI Adam BV and Others v Stichting de Thuiskopie and Stichting Onderhandelingen Thuiskopie vergoeding*, C-435/12, EU:C:2014:254

CJEU, *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others*, C-360/13, EU:C:2014:1195

CJEU, *Copydan Båndkopi v Nokia Danmark A/S*, C-463/12, EU:C:2015:144

CJEU, *GS Media BV v Sanoma Media Netherlands BV and Others*, C-160/15, EU:C:2016:644

CJEU, *Filmspeler*, C-527/15, EU:C:2017:300

CJEU, *OSA*, C-351/12, EU:C:2014:110

CJEU, *Spiegel Online*, C-516/17, EU:C:2019:625

CJEU, *Cofemel — Sociedade de Vestuário SA v G-Star Raw CV*, C-683/17, EU:C:2019:721

CJEU, *Brompton Bicycle Ltd v Chedech/Get2Get*, C-833/18, EU:C:2020:461

CJEU, *Constantin Film Verleih GmbH v YouTube LLC and Google LLC*, C-264/19, EU:C:2020:542

CJEU, *VG Bild-Kunst v Stiftung Preußischer Kulturbesitz*, C-392/19, EU:C:2021:181

CJEU, *Commissione c. Italia*, Case 39/72, EU:C:1973:13

CJEU, *Commissione c. Danimarca*, C-541/16, EU:C:2018:251

CJEU, Case C-250/25, *Like Company v Google Ireland Limited*, request for a preliminary ruling (pending; oral hearing 10 March 2026; AG Opinion expected 3 September 2026)

II. NATIONAL CASE LAW

District Court of Hamburg (Landgericht Hamburg), judgment of 27 September 2024, 310 O 227/23 (*LAION e.V. v Robert Kneschke*)

Landgericht München I, judgment of 11 November 2025, Az. 42 O 14139/24 (*GEMA v OpenAI*)

Getty Images (US) Inc and Others v Stability AI Ltd [2025] EWHC 38 (Ch)

TuneIn Inc v Warner Music UK Ltd [2021] EWCA Civ 441

BIBLIOGRAPHY

I. MONOGRAPHS AND CONTRIBUTIONS TO EDITED VOLUMES

Almada, M. (2025). The EU AI Act in a global perspective. In M. Furendal & A. Lundgren (Eds.), Handbook on the global governance of AI. Edward Elgar (forthcoming). Draft of 17 June 2025

Auteri, P., Floridia, G., Mangini, V. M., Olivieri, G., Ricolfi, M., Romano, R., Sironi, G. E., & Spada, P. (2023). Diritto industriale: Proprietà intellettuale e concorrenza (7th ed.). Giappichelli.

Benatti, F. Le 'nuove' eccezioni e limitazioni al diritto d'autore introdotte dal d.lgs. n. 177/2021. Giurisprudenza commerciale

Bertani, M. (2011). Diritto d'autore europeo. Giappichelli.

Bertani, M., Cogo, A., Fabbio, P., Genovese, A., & Ottolia, A. (2024). Lineamenti di diritto industriale: Concorrenza e proprietà intellettuale. CEDAM.

Burelli, C. (2026). Commento all'art. 1. In D'Amico, M. et al. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 471–481

Panigada, N. (2026). Commento all'art. 25. In D'Amico, M., Balaguer Callejón, F., Aguilar Calahorro, A., Gambatesa, P., Di Giovanni, S. (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132. Giappichelli, pp. 669–670

Rosati, E. (2021). Copyright in the Digital Single Market: Article-by-article commentary to the provisions of Directive 2019/790. Oxford University Press. https://doi.org/10.1093/oso/9780198858591.001.0001

Sganga, C. (2021). The right of reproduction. In The Routledge handbook of EU copyright law (pp. 123–150). Routledge

II. JOURNAL ARTICLES

Arnold, R., & Rosati, E. (2015). Are national courts the addressees of the InfoSoc three-step test? Journal of Intellectual Property Law & Practice, 10(10), 741–747

Bertani, M. (2006). Banche dati ed appropriazione delle informazioni. Europa e Diritto Privato, 2006.

Bogataj Jančič, M., & Purkart, E. (2025). Text and data mining in the Slovenian legal system. Stockholm Intellectual Property Law Review, 2024(2), 5–8

Cooper, A. F., & Grimmelmann, J. (2024). The files are in the computer: Copyright, memorization, and generative AI. arXiv preprint arXiv:2404.12590. https://arxiv.org/abs/2404.12590

Dornis, T. W., & Lucchi, N. (2025). Generative AI and the scope of EU copyright law: A doctrinal analysis in light of the referral in Like Company v. Google. IIC — International Review of Intellectual Property and Competition Law, 56, 1800–1840. https://doi.org/10.1007/s40319-025-01649-7

Geiger, C. (2007). From Berne to national law, via the Copyright Directive: the dangerous mutations of the three-step test. European Intellectual Property Review, 29(12), 486–491

Granieri, M. (2022). Il data mining nella disciplina del diritto d'autore e la strategia europea sui dati. AIDA, 2022

Hamann, H. (2024). Artificial intelligence and the law of machine-readability: A review of human-to-machine communication protocols and their (in)compatibility with Article 4(3) of the Copyright DSM Directive. JIPITEC — Journal of Intellectual Property, Information Technology and E-Commerce Law, 15(2)

Leistner, M., & Antoine, L. (2025). TDM and AI training in the European Union — From "LAION" to possible ways ahead? GRUR International, 74(11), 1027–1044. https://doi.org/10.1093/grurint/ikaf114

Lucchi, N., & Hunter, S. (2025). Generative AI and copyright — Training, creation, regulation (PE 774.095). European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs. https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)774095

Mansani, L. (2019). Le eccezioni per estrazione di testo e dati, didattica e conservazione del patrimonio culturale. AIDA, 2019.

Margoni, T. (2024). TDM and generative AI: Lawful access and opt-outs. Forthcoming in Auteurs & Media 2024. SSRN. https://ssrn.com/abstract=5036164

Quintais, J. P. (2025). Generative AI, copyright and the AI Act. Computer Law & Security Review, 56, 106107. https://doi.org/10.1016/j.clsr.2025.106107

Rosati, E. (2024). No step-free copyright exceptions: The role of the three-step in defining permitted uses of protected content (including TDM for AI-training purposes). European Intellectual Property Review, 46(5), 262–274

Servanzi, R. (2022). Le estrazioni di testo e di dati. Le Nuove Leggi Civili Commentate, 2022(5), 1150–1184.

III. INSTITUTIONAL REPORTS AND OPINIONS

European Copyright Society. (2025). Copyright and generative AI: Opinion of the European Copyright Society

EUIPO. (2025). The development of generative artificial intelligence from a copyright perspective. European Union Intellectual Property Office

Drexl, J. et al. (2021). Artificial Intelligence and Intellectual Property Law — Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate. Max Planck Institute for Innovation & Competition Research Paper No. 21-10

IV. TECHNOLOGICAL CONTRIBUTIONS (computer science)

Hans, A., Wen, Y., Jain, N., Kirchenbauer, J., Kazemi, H., Singhania, P., Singh, S., Somepalli, G., Geiping, J., Bhatele, A., & Goldstein, T. (2024). Be like a goldfish, don't memorize! Mitigating memorization in generative LLMs. In Proceedings of NeurIPS 2024

Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2022). Deduplicating training data makes language models

better. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)

Sakarvadia, M., Ajith, A., Khan, A., Hudson, N., Geniesse, C., Chard, K., Yang, Y., Foster, I., & Mahoney, M. W. (2024). Mitigating memorization in language models. arXiv preprint arXiv:2410.02159

Sugimura, P., & Hartl, F. (2018). Building a reproducible machine learning pipeline. arXiv preprint arXiv:1810.04570