Copyright and AI training

IRIS

A publication
of the European Audiovisual Observatory

European Audiovisual Observatory

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IRIS-2
Copyright and AI training
European Audiovisual Observatory, Strasbourg, 2026
ISSN 2079-1062

Director of publication – Pauline Durand-Vialle, Executive Director
Editorial supervision – Maja Cappello, Head of Department for Legal Information
European Audiovisual Observatory

Editorial team – Maja Cappello, Sophie Valais and Diego de la Vega
European Audiovisual Observatory

Author
Diego de la Vega

Proofreading
Michael Finn

Editorial Assistant – Alexandra Ross
Press and Public Relations – Alison Hindhaugh, alison.hindhaugh@coe.int
European Audiovisual Observatory

Publisher
European Audiovisual Observatory
76, allée de la Robertsau, 67000 Strasbourg, France
Tel.: +33 (0)3 90 21 60 00
iris.obs@coe.int
www.obs.coe.int

Cover layout – ALTRAN, France

Please quote this publication as
De la Vega, D., *Copyright and AI training*, IRIS, European Audiovisual Observatory, Strasbourg, July 2026

To promote inclusive language, we follow the *guidelines of the Council of Europe*.

Copyright and AI training

Diego de la Vega

Foreword

More than three centuries have passed since the Statute of Anne laid the foundations of modern copyright law by attempting to balance the rights of authors and editors. Its full title explained both the Statute's aim ("*the Encouragement of Learning*") and the means of achieving such a lofty objective ("*Vesting the Copies of Printed Books in the Authors or Purchasers of such Copies, during the Times therein mentioned*"). The first paragraph then outlined a problem presented as new but which, in reality, is as old as time: "[…] *Printers, booksellers, and other persons have of late frequently taken the liberty of printing, reprinting, and publishing […] books, and other writings, without the consent of the authors or proprietors […], to their very great detriment, and too often to the ruin of them and their families*".

The 1710 Statute of Anne centred on printed books and unsavoury publishers, whereas the *Loi du 13 janvier 1791 relative aux spectacles*, promoted by Pierre-Augustin Caron de Beaumarchais and other French playwrights, dealt with the unauthorised public performance of theatre plays. And each new era then brought its own technological developments, problems, and eventually, its own practical and legal solutions.

We are now in the era of artificial intelligence (AI), a revolutionary technological development which confronts copyright law with a new set of intricated problems. Problems that, ironically, echo the text of the Statute of Anne. Or at least, in one particular word of its title:

Learning.

Of course, the Statute of Anne aimed to encourage *people* to learn and was deeply concerned about inconsiderate profiteers who exploited *books and other writings* without *consent*. With it, a new (copy)right encouraged *learning* by securing a living for authors or proprietors who would otherwise have been dispossessed.

In the AI era, however, the abovementioned italicised words have somehow taken a different meaning. *Learning* is now conducted by machines, and the persons behind them insist that this very *learning* must take place *without the consent of the authors or proprietors*, probably without much concern for authors or proprietors' *very great detriment*.

Learning. Consent. Detriment.

This report discusses these three concepts by examining the legal issues arising during the training of AI systems. To this end, it explores how major copyright principles can be applied to the "input" phases of AI as well as the current interplay between copyright and generative and agentic AI. In doing so, it aims to provide a balanced and informed basis for understanding one of the defining topics in copyright policy not just in Europe, but also worldwide. A first (and maybe obvious) observation is that the approaches to copyright and AI are still very different depending on which side of the Atlantic we look at.

Enjoy the read!

Maja Cappello
IRIS Coordinator
Head of the Department for Legal Information
European Audiovisual Observatory

Table of contents

Figures

Tables

1. AI content generation in context. A copyright perspective

1.1. AI and content generation in Europe: the state of play

Since OpenAI released its public version of ChatGPT on 30 November 2022,[1] the use of artificial intelligence (AI) tools has increased at a very fast pace both for private and professional purposes. In January 2023, only two months after its release, the number of users of the platform had already risen to 100 million.[2] Many other AI platforms have emerged since then, offering users an enormous variety of functionalities.

The operation of these platforms, in particular when dealing with generative AI (GenAI) systems and assimilated technologies like Agentic, AI can be divided into two principal phases. The first is the training and the actual use of AI systems, and the second is the obtaining of the desired result which solves an actual problem through an output.

To analyse the copyright implications of these technologies, this report will look at the input phase of generative processes, like the training of AI systems and the prompting of instructions by users to obtain a specific output. Although AI training mainly concerns AI system providers and prompting mainly concerns users, under the current European framework both share significant implications in respect of copyright.

In this connection, AI utilities can fall within different categories of AI according to the results that users are seeking. For example, generative AI has recently been leading the use of AI, but research shows that the market focus is quickly shifting to other innovations such as AI agents (in the so-called "Agentic AI") that can perform tasks autonomously on behalf of users. On the other hand, mere GenAI is currently facing a temporary cycle of disillusionment.[3]

Nevertheless, only in the field of GenAI is an extensive range of tools also available today in a well-established market in Europe, where these services are normally

[1] OpenAI website, *Introducing ChatGPT*.
[2] Hu K., *ChatGPT sets record for fastest-growing user base - analyst note*, *Reuters,* 2 February 2023.
[3] Khandabattu H., *Latest hype cycle for artificial intelligence goes beyond GenAI*, *Gartner*, 8 July 2025. According to the different stages of the Gartner Hype, applicable to all emerging technologies, GenAI finds itself at the "Trough of Disillusionment" stage, where "interest wanes as experiments and implementations fail to deliver. Producers of the technology shake out or fail. Investments continue only if the surviving providers improve their products to the satisfaction of early adopters".

offered by large technological companies.[4] Around 20% of European companies and 55% of large enterprises[5] were already using AI in 2025 (mostly GenAI), representing an increase of 6.5% compared to 2024. It is also worth noting that roughly 10% of European businesses, such as production companies, are already using AI for the generation of pictures, videos and audio.[6] However, differences are emerging among individual territories. In terms of geographical distribution, statistics show that people in Northern Europe are making wider use of AI while southern countries are lagging behind. For example, 56.3% of Swedes are currently using GenAI in 2026 while 19.9% of Italians are using this technology.[7]

However, events like the launch of ChatGPT back in 2022 are just a milestone in AI. An important one to be sure, but just one more on a path that started back in the 1950s.[8] The constant evolution of AI results in continuous new functionalities by AI systems offered to users that reflect a very complex market environment. In a technology where the interest in its development is measured in "summers" and "winters",[9] authors believe that AI is currently going through a sustained summer, normally referred to as the "summer of AI". This is because of the growing, albeit somehow unstable, interest of companies and users in the development of AI as opposed to the past, when the interest in AI, in the context of broader technological development, was secondary.[10]

At the same time, legislators worldwide are attempting to provide an appropriate framework by addressing the various legal implications of AI. For example, the Council of Europe Framework Convention on Artificial Intelligence,[11] the first ever international legally binding treaty on AI, and relevant European legislation like the Artificial

[4] United Nations Conference on Trade and Development (UNCTAD), *Technology and Innovation Report 2025*, *UN-iLibrary*, Geneva, April 2025.

[5] Large enterprises in Europe are those employing 250 or more people. Eurostat website, *Statistics Explained* (consulted on 25 June 2026).

[6] Eurostat website, *Use of artificial intelligence in enterprises* (data extracted in December 2025).

[7] Eurostat, *The use of artificial intelligence (AI) technologies in the European Union*, Key Results, 2026 Edition.

[8] The foundations of AI can be found in Alan Turing's work "Computing machinery and intelligence", aiming to ascertain whether machines can think or not. Turing, a central figure in digital technology, also developed the so-called Turing Test, which tested the ability of machines to answer human questions, anticipating the chatbot model used nowadays by many popular AI platforms. The term "artificial intelligence" appeared as early as 1956, and critical experiments were already carried out throughout the 1960s. See Blagoj D., Chrisa T., Uroš K., *Historical Evolution of Artificial Intelligence*, EUR 30221 EN, *Publications Office of the European Union*, Luxembourg, 2020.

[9] Francesconi, E., *The winter, the summer and the summer dream of artificial intelligence in law*, Address to the 18th International Conference on Artificial Intelligence and Law. Artificial Intelligence and Law, *Springer Nature*, 3 February 2022.

[10] See Susskind, R. *The summer of AI* in *How to think about AI: a guide for the perplexed*, *Oxford Academic*. 20 March 2025. The author argues that events like the development of the World Wide Web meant a certain standstill in the development of AI.

[11] Council of Europe, *Framework Convention on Artificial Intelligence* was opened for signature on 5 September 2024 and aims to fill any legal gaps that may result from rapid technological advances from the point of view of human rights, democracy and the rule of law. It is technology-neutral and does not regulate technology itself, in an attempt not to be conditioned by regular developments in AI.

Intelligence Act (AIA)[12] have been introduced in the last few years. In parallel, as of June 2026, the European Commission (EC), is currently collecting evidence to assess the effectiveness of the existing legal framework and determine whether complementary measures are needed.[13] More specifically, according to the Call for Evidence on copyright, issued in May 2026, "the preparatory work for the targeted legislative initiative will run in parallel with the review of the CDSMD. This is why a single call for evidence is launched: it aims to collect the information necessary to support the review of the CDSMD, and to seek feedback on the challenges linked, in particular, to the application and enforcement of copyright and related rights in the context of technological developments and potential ways to address them".[14]

However, while general legislation is already in force, the position of the Court of Justice of the European Union (CJEU) on GenAI and copyright, meant to complement and interpret European legislation already in force, is yet to be known. In the field of copyright, its position is deemed important also because there is no specific copyright legislation for GenAI. In this area, copyright rules still fall within the scope of the general copyright framework.

In this connection, the CJEU's decision in the *Like Company v. Google Ireland Limited* case[15] is expected to be the first ever ruling in this matter at EU level. In connection with the protection of press publications concerning online uses,[16] it will presumably clarify some of the critical aspects of the existing European copyright regulation with respect to AI. While an overarching case has still to come to harmonise existing rules, other rulings like the recent *Meta Platforms Ireland* case,[17] can indirectly provide some guidance on critical aspects of AI and copyright, such as the role of copyright exclusive rights in the context of the CDSMD, which cannot be substituted for

---

[12] *Regulation (EU) 2024/1689* of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act - AIA).

[13] *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 (the CDMSD).

[14] European Commission, *Call for Evidence*, Report on the review of the Copyright in the Digital Single Market Directive / Targeted initiative for a better copyright environment for European creativity and innovation, Section A, 13 May 2026.

[15] European Union, *Case C-250/25*, Like Company, Request for a preliminary ruling from the Budapest Környéki Törvényszék (Hungary) lodged on 3 April 2025 – *Like Company v. Google Ireland Limited*, *EUR-Lex,* 16 June 2025. The request for a preliminary ruling, stemming from the Budapest *Környéki Törvényszék* (Hungary), seeks guidance *inter alia* on topics like the consideration of the responses by LLM-based AI chatbots as communication to the public, the consideration of the process of training an LLM-based chatbot as an instance of reproduction, the overall role of text and data mining exceptions in AI and certain prompting techniques.

[16] Article 15 CDSMD.

[17] Court of Justice of the European Union, *Meta Platforms Ireland Ltd v. Autorità per le Garanzie nelle Comunicazioni* [GC], Case C-797/23, 12 May 2026. In paragraph 62 thereof the Grand Chamber establishes that "Member States cannot transpose Article 15 of Directive 2019/790 by substituting the exclusive rights of a preventive nature which it establishes with a mere right to compensation, which would allow publishers of press publications only to obtain remuneration for online uses of those publications by information society service providers, but not to prohibit those uses".

the right to compensation for press publishers. This is also expected to provide guidance to the various national courts that are already delivering decisions on cases involving AI and copyright.

1.2. AI functionalities and the process of content generation

As acknowledged by the AIA, AI is a fast-evolving family of technologies. According to the Regulation, these technologies can ultimately represent a wide array of economic, environmental and societal benefits across the entire spectrum of industries and social activities. Their training is based on the premise of analysing big datasets which can be subject to copyright.

More specifically, these technologies can contribute to creating content, to enhancing prediction processes and to optimising all sorts of daily-life operations and resource allocation, and they can also personalise digital solutions available for individuals and organisations.[18] These functionalities are offered either on a stand-alone basis or as a component of a given digital product[19].

In the field of GenAI, different functionalities can be identified depending on the problem solved using this technology. GenAI virtual assistants are one of the most common functionalities and are based on large language models (LLMs) in order to generate high-quality outputs. They are followed by other AI system models like multimodal GenAI, which has the ability to deal with a blended mix of inputs like images, video or audio, generating complex outputs text-to-image or script-to-video content.[20]

The more recent Agentic AI can also be classified as a type of GenAI where systems not only generate content but also make autonomous decisions and take specific action related to them without an entirely explicit input. AI agents can achieve complex goals and design and execute workflows with different degrees of human intervention. Within the next three to six years, Agentic AI is expected to undergo a period of major development.

Finally, among the broad group of developing GenAI technologies, AI code assistants, capable of creating software code (i.e. for app or video game development),

[18] Recital 6 AIA. This content can also include, for example, deep fakes, in connection to personality and image rights, defined as AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful. Other implications can be, for example, freedom of expression. See the Council of Europe's *Guidance Note on the implications of Generative AI for freedom of expression*, adopted by the Steering Committee on Media and Information Society (CDMSI) in December 2025.

[19] For example, *Adobe's GenAI and Agentic AI* systems are "embedded across Adobe's creative and marketing applications", *Microsoft's Copilot* is integrated into the "Microsoft 365 ecosystem" while some of the most common AI tools are stand-alone chatbots like *ChatGPT* or *Claude*.

[20] Zimmermann, A. et al., *Emerging Tech Impact Radar: Generative AI*, *Gartner*, 14 February 2025.

with a great degree of autonomy and simple prompting, will presumably also develop at a very fast pace.[21]

In each economic sector, the impact of AI can be measured differently. In some economic sectors the impact and role of copyright can be crucial, while in others the impact may still be limited.[22] In those where GenAI can play an important role, like the audiovisual sector,[23] especially in purely creative tasks,[24] or video games development,[25] the impact of AI already tends to be high. As a rule, knowledge-intensive sectors where cognitive tasks are required are more likely to have a high exposure to AI. In those sectors, GenAI and synthetic content creation are expected to expand the effects of already non-GenAI uses.[26]

1.3. A technological approach to AI content generation

GenAI is a subtype of AI. According to all the tasks that can be carried out by GenAI, it can technically be defined as a type of AI based on prompting and the use of natural language that is based on deep learning (DL), which is a subtype of machine learning (ML) that can accomplish sophisticated tasks by using extensive datasets and appropriate training with the aim of creating new content.[27]

From a legal point of view, the European Parliament has recently adopted a resolution on Copyright and generative artificial intelligence (so-called "Voss" Report), where it defines GenAI as "a type of AI that, unlike other AI systems designed to primarily classify or predict, generates new content, such as text, images, music, videos and code, on the basis of training using very large datasets from which they learn patterns and structures". To this end, GenAI systems use predictions based on statistical models, can mimic human creativity and rely on pre-existing content.[28]

---

[21] Ibid.

[22] Eurostat, *The use of artificial intelligence (AI) technologies in the European Union*, op. cit. Eurostat. For example, sectors like construction are not as affected by AI as others like creative industries, where AI, and especially AI, can undertake a vast number of tasks.

[23] For a specific overview of AI in the audiovisual sector, see Cappello M. (ed.), *AI and the audiovisual sector: navigating the current legal landscape*, IRIS, European Audiovisual Observatory, Strasbourg, October 2024.

[24] For example, see Netflix's guidelines for *Using Generative AI in Content Production*, addressed to partner companies producing audiovisual content for Netflix. In prompting, the guidelines encourage avoid using inputs (e.g., prompts, images) that reference copyrighted materials or likenesses of public figures or deceased individuals without appropriate permissions.

[25] Heinisch, O. and Jadot, C. *Some Implications of the EU AI Act on Video Game Developers*, *Sheppard*, 28 February 2025.

[26] OECD, *Macroeconomic productivity gains from Artificial Intelligence in G7 economies*, *OECD Artificial Intelligence Papers*, June 2025, No. 41.

[27] See Microsoft, *Generative AI vs. other AI types*..

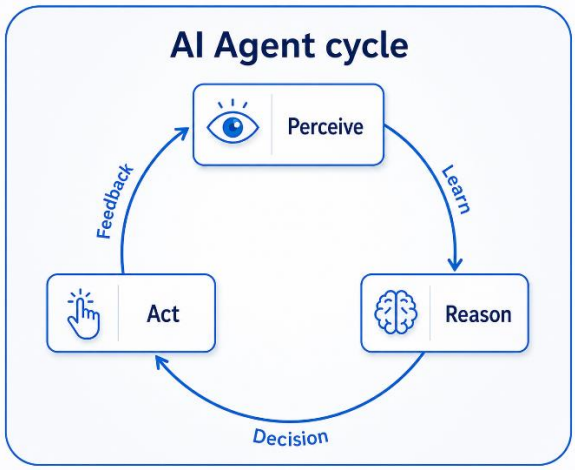
[28] European parliament, *Copyright and generative artificial intelligence – opportunities and challenges*, resolution of 10 March 2026 on copyright and generative artificial intelligence – opportunities and challenges (2025/2058(INI)) (Voss Report).

More broadly, the AIA, refers to general purpose AI models as allowing for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks.[29]

Other classifications of AI technology, like Agentic AI, remain, for the moment, a commercial and technical matter, without a legal approach[30] to their definition in Europe. Within the industry, AI agents are defined as AI systems that can encompass a wide range of functions beyond natural language processing including decision-making, problem-solving, interacting with external environments and performing actions.[31]

Beyond Europe, Australia has just defined AI agents as a "software-based system that perceives inputs from its environment, maintains or updates an internal state, and selects and executes actions using APIs or other tools to achieve defined goals within assigned permissions and constraints. It may plan, learn from feedback when permitted, and self-monitor its behaviour to improve performance over time, while operating within defined governance, oversight, and control mechanisms".[32] AI agents can eventually generate content too.

Figure 1. The Agentic AI process



Source: Australian Government

In brief, the underlying technical concepts are similar to those in general AI, where machine learning plays a fundamental role. Machine learning is based on the learning of preexisting datasets by the use of algorithms and the inferring of new data. A part of it is

[29] Recital 99 AIA.

[30] From an informative point of view, the European Data Protection Supervisor defines Agentic AI as "a concept in artificial intelligence (AI) that describes systems acting autonomously with limited human interactions (in particular, without step-by-step instructions) to fulfil goals rather than isolated tasks". Goldstein A., *Agentic AI*, European Data Protection Supervisor.

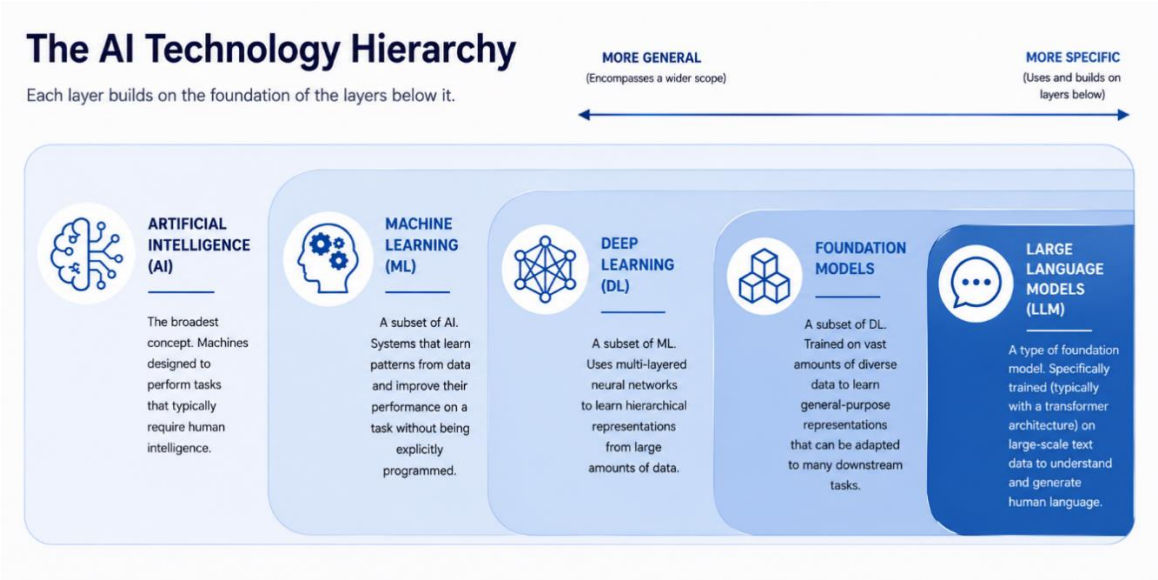
[31] Gutowska A., *AI agents defined*, *Think*, IBM.

[32] Australian Government (Digital Transformation Agency), *Agentic AI addendum to the AI technical standard for Australian Government*, *digital.gov.au* (last updated on 4 June 2026.

deep learning, which, according to the industry, is an "intricate series of nested equations that maps an input to an output".[33] Finally, foundation models are specific deep learning models which are trained on big amounts of available data,[34] from which LLMs have the ability to interact conversationally with people.[35]

Legislation regulating AI, be it from the copyright point of view or a broader perspective, assumes that the technical layers of AI are always based on the aforementioned principles. However, new technologies that start to emerge, like Quantum AI,[36] could challenge this assumption in the future as the way AI systems are trained can be significantly different when quantum computation and AI become entangled.[37]

Figure 2. Summary of basic underlying technologies in AI and GenAI



Source: European Audiovisual Observatory

[33] Bergmann, D., *What is deep learning?*, *Think*, IBM.
[34] See *Defining AI and chatbots*, Stanford Teaching Commons website, Standford University (consulted on 16 June 2026). See also Bergmann, D., *What is deep learning?*, op. cit.; Cappello M. (ed.), *AI and the audiovisual sector: navigating the current legal landscape*, op. cit.
[35] Jurafsky, D. and H. Martin, J., *Speech and Language Processing* (3rd ed. Draft), *Standford University*, Chapter 7.
[36] As described by Amazon Web Services, Quantum AI is the use of quantum technologies for running AI systems, as opposed to classical AI, based on ordinary computing. This technology would hypothetically replace current technical standards and enhance the working of AI systems and platforms. See Amazon Web Services (aws) website, *What is Quantum AI?* (consulted on 10 June 2026).
[37] Acampora G., Chiatto A. et al. *Quantum artificial intelligence: a survey*, *Computer Science Review 59*, February 2026.

1.4. The interplay between copyright and AI

Since GenAI has the ability to create content, copyright issues already arise already in the case of its actual generation process. One of the early concerns about the use of GenAI since 2022 and the irruption of GenAI tools onto the market has been, in particular, the use of preexisting content for the massive training of AI systems. For example, as early as 2023, the scriptwriters' strike in the USA showed that, insofar as AI tools are used in the audiovisual industry, copyright-related disputes between producers and authors are likely to quickly emerge.[38]

In the context of the European Democracy Shield and the media sector, the EU institutions have jointly warned about the impact of AI training. While praising the development of AI, EU institutions warn that "the use of copyrighted material to train AI models without permission remains a significant threat to the media sector, eroding revenues and impacting on quality and diversity of media".[39]

As expressed in the previously mentioned European Parliament "Voss Report", AI training generally "may include copyright-protected materials". The Report warns that "the development, deployment and use of AI must be fully compliant with the current legal framework". It depicts a gloomy atmosphere globally, especially for the EU, where there is "evidence of the widespread violation of copyright rules by GenAI providers, including the unauthorised collection of works from the internet" as well as critical problems like non-compliance with rightsholders' TDM rights reservations, use of pirated sources to obtain works, and the failure to seek licences to use preexisting works.[40]

The Parliamentary Assembly of the Council of Europe (PACE) has also recently warned about similar concerns. According to PACE, "the emergence of the artificial intelligence era has given rise to a particularly challenging set of problems for the creative sector because in order to feed their data-hungry systems, AI companies are scraping the internet without prior permission and without remunerating content creators on the basis of legislative provisions that are neither clear-cut nor fit for purpose".[41]

Legally, this situation does indeed require the application of copyright law. The training of AI systems typically involves, among other acts and techniques, massive acts of reproduction and eventually the extraction of content from databases[42]. It is closely connected to text and data mining (TDM), which enables the processing of large amounts

---

[38] Cullins A. and Kilkenny K., *As Writers Strike, AI Could Covertly Cross the Picket Line*, *The Hollywood Reporter*, 3 May 2023.

[39] European Commission, *Joint Communication to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions. European Democracy Shield: Empowering Strong and Resilient Democracies*, JOIN(2025) 791 final, Brussels, 12 November 2025.

[40] Recital L Voss Report.

[41] Parliamentary Assembly, *Resolution 2654 (2026)*, Copyright enforcement in the artificial intelligence environment, adopted on 23 April 2026.

[42] Some online platforms offer users the possibility of partly tracing the use of preexisting material in the training of AI systems. For example, the *AI Watchdog* from The Atlantic offers an "investigation of the books, videos, and other media used by the world's most powerful tech companies to train their AI models".

of information with a view to gaining new knowledge and discovering new trends[43] in order to obtain specific information useful for the training.[44]. In particular, it is still unclear whether the provision of an input in the form of a prompt given by the user, if based on copyrighted content fed during the training process, such as press publications, entails an act of reproduction of copyrighted material.[45]

The balance between all the rights and interests at play has recently been summarised by the British Government's Copyright and AI Impact Assessment. It acknowledges the value of copyright law in sustaining the success of the creative industries as it "enables creators to earn recognition and financial benefit from their work, while attracting investment for the future". However, as explicitly recognised, "rightsholders say they face challenges in controlling use of and access to their works. AI developers say they struggle to access large quantities of data to train foundation models [...]".[46]

Copyright regulation therefore plays a fundamental role at the intersection of the interests of rightsholders and those of AI developers, and the platform industry originated around the provision of AI services. GenAI, Agentic AI and other assimilated technologies are the main concern of legislators as their effects on AI training, and to a lesser extent, prompting can have a great impact not only in terms of legal certainty, but also economically and societally.

[43] Recital 8 CDSMD.
[44] Rosati, E., *The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law*, *Computer Law & Security Review*, Vol. 59 in *Science direct*, November 2025.
[45] European Union, *Case C-250/25,* op. cit.
[46] HM Government, *Copyright and AI Impact Assessment*, Presented to Parliament pursuant to Section 135 of the Data (Use and Access) Act 2025. Department for Science, Innovation and Technology and Department for Culture, Media and Sport, March 2026.

2. The role of text and data mining exceptions in AI training

2.1. Regulating AI: General European AI framework

The Special Eurobarometer on the Digital Decade 2026 shows that 80% of Europeans favour the careful regulation of AI to ensure safety over development without restrictions.[47] This feeling among Europeans is aligned with the rationale of the AIA for which, "given the major impact that AI can have on society and the need to build trust, it is vital for AI and its regulatory framework to be developed in accordance with Union values".[48] At a broader level, the objectives of the Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law are also "to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law".[49]

In the EU, the AIA must be understood as part of a global strategy that goes beyond regulation, encompassing innovation plans and a broader strategy at policy level.[50]

The AIA, still in the process of being fully applicable,[51] remains today as the fundamental AI regulatory tool in Europe. Rather than on copyright, it focuses on the functioning of the internal market and the protection of health, safety and fundamental

[47] European Commission: Directorate-General for Communication and Directorate-General for Communications Networks, Content and Technology, *Digital decade – Eurobarometer report – February – March 2026*, Directorate-General for Communications Networks, Content and Technology, 2026. The result shows territorial differences. While citizens in Greece, Luxembourg and Ireland favour AI regulation, people in Romania, Poland and Estonia show a lower level of support, even if a majority is still in favour of regulation over a more permissive approach. The report also shows that women are more likely than men to favour AI regulation and that, generally speaking, younger Europeans are less in favour of regulation. See pages 60 and 61.

[48] Recital 6 AIA.

[49] Council of Europe, *Framework Convention on Artificial Intelligence*, Article 1.

[50] European Commission website, *European approach to artificial intelligence* (consulted on 1 June 2026).

[51] The AIA is applicable as of 2 August 2026, although some chapters have been in force since 2025 and Article 6.1 (regarding Classification rules for high-risk AI systems) is to apply from 2 August 2027.

rights.[52] At the national level, several jurisdictions have either AI laws in force (Denmark, Finland, Italy and Slovenia) or legislative proposals in place.[53]

However, when it comes to copyright law, apart from some specific parts of the AIA that refer to general-purpose AI systems like Article 53, the AI points directly to the CDSMD. For example, Recitals 105 and 106 of the AIA expressly refer to the CDSMD, as it "introduced exceptions and limitations allowing reproductions and extractions of works or other subject matter, for the purpose of TDM, under certain conditions".[54] Therefore, the AIA confers on the CDSMD significant powers in AI regulation matters related to copyright.

2.2. Copyright framework

The CDSMD is one of the latest pieces of the complex European copyright framework.[55] It complements a number of existing Directives which contain the main provisions upon which national copyright systems are based, especially Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society ("InfoSoc" Directive).[56] It introduced as early as 2001 some of the main architectural elements that still hold up the European copyright framework, and the CDSMD builds on them.

The CDSMD is the central piece of copyright regulation in connection with AI, as there are no specific rules addressing copyright in the AIA; its rationale follows the line of all the rest of the preceding copyright Directives.[57]

2.2.1. The European copyright framework and the rules applicable to AI training

The CDSMD predates the dawn of GenAI around 2022-2023[58] and contains no mention of AI while acknowledging that "new technologies enable the automated computational

[52] Article 1 AIA.
[53] Bashir, E., *Status of national AI laws across Europe*, *Cullen International*, 12 June 2026.
[54] Recital 105 AIA.
[55] This general framework applies to all sectors. For example, in the audiovisual sector, it can apply to assisted script writing, AI-assisted video editing, subtitling and dubbing, prompted video generation, content recommendation and much more.
[56] *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 (the InfoSoc Directive).
[57] Apart from the InfoSoc Directive, they include the following (which can have an impact on AI) are: Directives *96/9/EC* of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *2006/115/EC* on rental right and lending right and on certain rights related to copyright in the field of intellectual property, and *2009/24/EC* on the legal protection of computer programs.

analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining" also referring to the "development of new applications or technologies".[59] Also, according to Recital 3 CDSMD, "relevant legislation needs to be future-proof so as not to restrict technological development". Immediately after the CDSMD was adopted, the European Parliament stated that "furthermore, in addition to the mandatory exception for text and data mining for research purposes proposed by the Commission in its proposal, the co-legislators agreed to enshrine in EU law another mandatory exception for general text and data mining (Article 4) in order to contribute to the development of data analytics and artificial intelligence and on a number of new EU rules (e.g. new licensing mechanism for out-of-commerce works)".[60]

As the recent Call for Evidence from the European Commission for the review of the CDSMD acknowledges, "the CDSM Directive and the copyright-related provisions of the AI Act and subsequent implementation measures, contribute to supporting rightsholders in the exercise of their rights. However, they may not be sufficient to address the difficulties faced by rightsholders in controlling and licensing their works for AI uses".[61] For this reason, the EU institutions already announced in late 2025 that the Commission would review the CDSMD and consider how to improve its effectiveness in the context of GenAI. Notably, as stated in the Joint Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions "the development of generative AI, while bringing opportunities, presents various challenges to the media ecosystem. Online piracy and the use of copyrighted material to train AI models without permission remains a significant threat to the media sector, eroding revenues and impacting on quality and diversity of media. The Commission will review the Directive on copyright in the Digital Single Market and consider how to improve its effectiveness in the context of these developments".[62]

For the application of copyright to take place, it is essential that pre-existing works must fall within the scope of national copyright laws, meaning that some of them might not be relevant in terms of AI training. For example, works in the public domain (although moral rights may subsist) or content not considered a work in terms of

---

[58] The first proposal of the European Commission dates to 14 September 2016: *COM(2016) 593 final* on copyright in the Digital Single Market. Preparatory studies used for the CDSMD, which are even older, also do not mention AI. See Triaille J-P., de Meeûs, J. and de Francquen A., *Study on the legal framework of text and data mining (TDM)*, *Publication office of the European Union*, March 2014.

[59] Recitals 8 and 18 CDSMD.

[60] European Parliament, *Legislative Train Schedule*, Modernisation of European copyright rules: directive on copyright in the digital single market (information updated as of 20 November 2011).

[61] European Commission, *Call for Evidence for an impact assessment*, Targeted initiative for a better copyright environment for European creativity and innovation.

[62] *Joint Communication to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions. European Democracy Shield: Empowering Strong and Resilient Democracies*, op. cit.

copyright are not fully protected.[63] In this sense, general copyright provisions on work protectability apply.

2.2.1.1. Rights involved

The following rights can be identified according to the procedures deployed by AI model providers in the process of training that may affect the copyright in preexisting works and therefore prevent AI training or parts thereof:[64]

Table 1. Copyright involved in the training of AI systems

Area		Connection to AI training
	**Reproduction Right**	The reproduction right entitles rightsholders to an exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part, of i.e. works, phonograms and performances.[65]
	**Database Rights**	Databases enjoy two kinds of protection: - Copyright protection: it introduces exclusive rights to authorise reproductions of their databases, irrespective of whether such reproductions are temporary or permanent.[66] - *Sui generis* right: it introduces the right to prevent extraction and/or re-utilisation (of the whole or of a substantial part, evaluated qualitatively and/or quantitatively) of the contents of that database.[67]
	**Computer Program Rights**	Building on standard copyright protection stemming from the InfoSoc Directive, there is a specific right of reproduction either permanent or temporary and alteration which includes translation, adaptation, and arrangement.[68]
	**Press Publisher's Rights**	The CDSMD provides for specific protection in the form of a neighbouring right against certain online uses by information society service providers. Recent interpretation from the CJEU allows national legislation, inter alia, to confer on publishers of press publications the right to obtain fair remuneration in

[63] As explained in Recital 53 of the CDSMD, the expiry of the term of protection of a work entails the entry of that work into the public domain and the expiry of the rights that Union copyright law provides in relation to that work.
[64] *The development of Generative Artificial Intelligence from a Copyright Perspective*, op. cit., p. 35.
[65] Article 2 *InfoSoc Directive*.
[66] Article 3 Directive *96/9/EC*.
[67] Article 7 Directive *96/9/EC*
[68] See Directive *2009/24/EC*.

Area		Connection to AI training
		return for the authorisation to use their publications granted to information society service providers provided that: - they do not deprive publishers of press publications of the possibility of refusing to grant such authorisation or that of granting it free of charge; - that it does not impose on information society service providers any payment obligation unrelated to the use of such publications; - that the obligations and any penalties imposed on those providers observe the principle of proportionality.[69]
	**Technological Protection Measures (TPM)**	The InfoSoc Directive requires national laws of EU jurisdictions to provide adequate legal protection against the circumvention of effective technological measures used by rightsholders, as long as they do not prevent the beneficiaries of specific limitations and exceptions to copyright and related rights to benefit from them.

Source: European Audiovisual Observatory and EUIPO[70]

2.2.1.2. Relevant exceptions

Copyright in Europe is counterbalanced by a set of exceptions, however. This concept, enshrined in particular in the InfoSoc Directive, aims to ensure a fair balance of rights and interests between the different categories of rightsholders in the copyright environment. In line with this reasoning, the CDSMD introduced new exceptions and limitations for text and data mining, teaching and preservation of cultural heritage.[71]

From the belief that certain organisations and institutions come up against legal uncertainty as to the extent to which they can perform TDM of content, the CDSMD introduced the TDM exception, also with a view to ensuring that this technique, when carried outside the scope of scientific research, is lawful.

Accordingly, when the use of TDM techniques is carried out by both private and public entities to analyse large amounts of data in different areas of life and for various general purposes, the CDSMD introduces an exception for TDM when the work or other subject matter is accessed lawfully by the beneficiary. This lawful access includes when it has been made available to the public online, and insofar as the rightsholders have not reserved in an appropriate manner the rights to make reproductions and extractions for text and data mining.[72]

[69] See *Case C-797/23*, op. cit.
[70] *The development of Generative Artificial Intelligence from a Copyright Perspective*, op. cit.
[71] Recital 5 CDSMD.
[72] Recital 18 CDSMD.

In other words, the CDSMD introduces the "opt-out regime", which has rapidly become one of the most relevant topics of AI training. Some voices wonder about its application to the training of AI systems[73] and its actual effects in the market. These questions persist seven years after its introduction and four years after the popularisation of GenAI among the general public.[74]

In this connection, TDM is dealt with in Articles 3 and 4 CDSMD. Article 4 is actually the core of the policy discussion since it provides for a broad exception while providing for an opt-out regime and requiring rights holders to exercise the opt-out through machine readable means in case of content publicly available:

Table 2. TDM exceptions relevant in the AI training process

CDSMD Article		Exception
	**3**	Exception for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.
	**4**	Exception or limitation for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining. - The exception or limitation shall apply on condition that the use of works and other subject matter referred to in that paragraph **has not been expressly reserved** by their rightsholders in an appropriate manner, **such as machine-readable means** in the case of content made publicly available online [opt-out regime].

Source: European Audiovisual Observatory

[73] Dornis, T., *The Training of Generative AI is not Text and Data Mining*, European Intellectual Property Review (E.I.P.R.). 19 October 2024. The author argues that the focus is rather on copyright infringement since the training of generative AI models without licences for the works used as training data does not actually fall within the scope of the European TDM exception. See also *Copyright and generative artificial intelligence opportunities and challenges*, European Parliament resolution of 10 March 2026 on copyright and generative artificial intelligence – opportunities and challenges (2025/2058(INI)), paragraph K.

[74] Keller, P. (IViR) *A Blind Spot at the Heart of EU Copyright and AI Policymaking?*, *Kluwer Copyright Blog*, 27 May 2026. The author believes that apart from the uncertainty around the implementation of the opt-out regime, it fails to generate the expected results because "the incentives around the mass-scale use of publicly available information and content are shaped by dynamics that extend far beyond the reach of the EU copyright framework".

2.2.2. Text and data mining and transparency obligations

Text and data mining is defined by the CDSMD as any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.[75]

With regard to copyright, text and data mining can involve acts protected either by copyright or by the *sui generis* database right, or occasionally by both. In particular, it can happen in the reproduction of works (or also other subject matter), like when the extraction of contents from a database occurs when the data are normalised[76] in the process of text and data mining.[77] This results in the automated computational analysis of information in digital form, such as text, sounds, images or data, that generally falls within the scope of TDM.

Six years after the introduction of the CDSMD, according to the "Voss Report", the current systems for the reservation of rights to copyright-protected content (the CDSMD states that "it should only be considered appropriate to reserve those rights by the use of machine-readable means"[78]) "are often impractical, may not cover all relevant acts of text and data mining, and lack the necessary transparency for effective implementation and enforcement". The report also acknowledges that there is evidence of the widespread violation of copyright rules by GenAI providers, including non-compliance with rightsholders' text and data mining rights reservations.[79] PACE also stresses that "the so-called 'text and data mining (TDM) exceptions', which were adopted before the advent of generative AI, place the onus on copyright holders to opt out of this exception and do not include any remuneration provisions".[80] Nevertheless, the opt-out regime, as connected to exclusive rights, allows for the existence of subsequent licensing and remuneration schemes.

Meanwhile, in 2025 the European Commission made public the General-Purpose AI Code of Practice,[81] signed by some of the most relevant AI developers or providers (including Amazon, Anthropic, Google, IBM, Microsoft or Mistral AI). The signatories to it acknowledged[82] the exception or limitation for text and data mining "which shall apply on conditions of lawful access and that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightsholders in an appropriate manner […]".

According to this Code, the signatories commit to ensuring that only lawfully accessible works and other protected subject matter are reproduced and extracted. To

[75] Article 2(2) CDSMD.
[76] Organised, cleaned and standardised.
[77] Recital 8 CDSMD.
[78] Recital 18 CDSMD.
[79] Section K of Voss Report.
[80] Recital 6 of *Resolution 2654 (2026)*, Copyright enforcement in the artificial intelligence environment, op. cit.
[81] European Commission, *General-Purpose AI Code of Practice now available*, press release, 10 July 2025.
[82] Idem, Copyright Chapter.

that end, they have adopted two commitments governing the use of web-crawlers to scrape or compile data for the purpose of text and data mining:

- **Not to circumvent effective technological measures** as defined in the InfoSoc Directive that are designed to prevent or restrict unauthorised acts in respect of works and other protected subject matter, in particular by respecting any technological denial or restriction of access imposed by subscription models or paywalls.
- To **exclude from their web-crawling websites** that make content available to the public and are, at the time of web-crawling, recognised as persistently and repeatedly infringing copyright and related rights on a commercial scale by courts or public authorities in the EU and the European Economic Area.

At the same time, in order to identify and comply with rights reservations when crawling the Internet, in line with the spirit of Article 4 CDSMD, two further commitments were adopted by the signatories from a technical point of view:

- To employ web-crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt), as specified in the Internet Engineering Task Force (IETF) Request for Comments No. 9309.[83]
- To **identify and comply with other appropriate machine-readable protocols** to express rights reservations pursuant to Article 4(3) CDSMD. This can be done, 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. These may include those that are technically implementable, and widely adopted by rightsholders, considering different cultural sectors, and generally agreed through an inclusive process based on *bona fide* discussions to be facilitated at EU level.

With regard to transparency in GenAI training, the AIA introduces, inter alia, an obligation for providers of AI systems, including general-purpose AI systems generating synthetic audio, image, video or text content, to draw up and keep up-to-date the technical documentation of the model, including its training and testing process and the results of its evaluation.[84] This documentation must include the modality (e.g. text, image) and format of inputs and outputs, the design specifications of the model and training process, including training methodologies and techniques, the key design choices including and the rationale and assumptions made.[85] Also, Article 53(1)d CDSMD obliges providers of general-purpose AI models to draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model.

Along with these, providers must also clarify the information on the data used for training, testing and validation, where applicable, including the type and provenance of data and curation methodologies (e.g. cleaning, filtering, etc.), the number of data points, their scope and main characteristics; how the data was obtained and selected as well as

[83] See Datatracker website, *Robots Exclusion Protocol RFC 9309*, op. cit.
[84] Article 53.1.a. AIA.
[85] Annex XI AIA.

all other measures to detect the unsuitability of data sources and methods to detect identifiable biases, where applicable.[86]

As provided for in Articles 53(2) and 54(6) AIA and clarified by the Commission Guidelines on the scope of the obligations for providers of general-purpose AI models established by Regulation (EU) 2024/1689, special exceptions apply to providers of AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model.[87]

2.2.3. Selected examples of approaches to copyright protection in AI

Copyright regimes precede the AI era (as other historical technological developments), confronting legislators with the need to ascertain whether European and national copyright frameworks are fit for the GenAI era or not.

For example, the recent Copyright and AI Impact Assessment of the UK[88] summarises the different approaches to copyright legislation that can be identified in today's technological context and considering major developments like the EU copyright framework of the CDSMD.

In the UK, there is currently no broad exception that can underpin the use of copyrighted works for commercial TDM or the training of GenAI.[89] In discussing future policy, the British assessment has marked three objectives, namely increasing transparency about the use of copyright works to train AI models and AI-generated content, giving rightsholders of creative works greater control over the use of their material to train AI models and supporting their ability to be remunerated where it is used; and finally, enhancing the ability of AI developers to access the material they need to train world-leading AI models in the UK. According to this set of objectives, four options typically arise:

[86] Ibid.
[87] European Commission, *Guidelines on the scope of the obligations for providers of general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 7719 final, Brussels, 19 November 2025.
[88] HM Government, *Copyright and Artificial Intelligence: Impact Assessment*, Department for Science, Innovation and Technology. Department for Culture, Media and Sport, presented to Parliament pursuant to Section 135 of the Data (Use and Access) Act 2025, 18 March 2026.
[89] UK parliament publications, *AI, copyright and the creative industries*, HL Paper 267, Communications and Digital Committee, 4th Report of Session 2024-2026, 6 March 2026.

Table 3. AI training. Policy alternatives

Option	Policy
0	**Do nothing**. Under this option, permission would usually be needed to copy protected works at different stages of AI training and development.
1	**Strengthen copyright**, so that licences are needed for uses of copyright material in AI training with transparency requirements and market access measures. In the UK, since copyright law already provides some rights, the licensing scheme under this option would remain largely the same as the *status quo*.
2	**Introduce a broad copyright exception for TDM**. The exception would not be entirely without constraint, as data mining activity would only be allowed where the user has lawful access. However, rightsholders would not be able to reserve their rights and "opt out" of the exception.
3	**Introduce a TDM copyright exception with a rights reservation mechanism** and transparency requirements. The exception would apply only if a rightsholder had not expressly reserved their rights. This reservation would allow rightsholders to "switch off" the exception on selected works.

Source: HM Government, Copyright and AI: Impact Assessment[90]

As copyright regimes vary worldwide, and the approaches to AI training differ depending on the jurisdiction, there are different situations worldwide with regard to the treatment of AI training in copyright policy. For example, in countries like France discussions are ongoing on the possible introduction of a presumption of the use of works to increase transparency on the use of copyright protected content in the context of AI.[91]

Unlike general instruments to harmonise AI legislation that are starting to emerge, such as the Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, there is no major convention at world level dealing with AI and copyright. Copyright continues to be governed, generally, by the Berne Convention,[92] other international treaties like the WCT[93] and the WPPT,[94] and national (and European) instruments.

As of 2026, these are some of the current approaches to AI and copyright:

[90] *Copyright and AI Impact Assessment*, op. cit., p. 15.
[91] See the *Sénat* website, *Proposition de loi relative à l'instauration d'une présomption d'exploitation des contenus culturels par les fournisseurs d'intelligence artificielle*, adopted on 8 April 2026 (consulted on 23 June 2026).
[92] World Intellectual Property Office (WIPO), *Berne Convention for the Protection of Literary and Artistic Works*, adopted on 9 September 1886.
[93] WIPO, *WIPO Copyright Treaty (WCT)*, Geneva, adopted on 20 December 1996.
[94] WIPO, *WIPO Performances and Phonograms Treaty*, Geneva, adopted on 20 December 1996.

Table 4. European and national approaches to AI training

	**EU**	TDM exceptions provided in Articles 3 and 4 CDSMD. Only regime including an "opt-out" mechanism.
	**UK**	Statutory exception permitting TDM for non-commercial research only, provided the user has lawful access to the work. Policy discussions being held.
	**USA**	No specific tool for TDM. Fair use doctrine applies.[95]
	**CANADA**	No specific tool for TDM. Policy discussions being held.[96]
	**JAPAN**	Broad statutory exception allowing TDM for commercial and non-commercial uses, particularly for "non-enjoyment purposes".[97]
	**CHINA**	No specific framework. Copyright regime is a mix of the copyright system (including fair use) and authorship system.[98]

Source: European Audiovisual Observatory

In the case of European rules, it must be taken into account that some of the territorial scope rules[99] set out in the AIA establish that the AIA applies to deployers of AI systems

[95] The American approach to AI training is especially important considering that some of the major AI providers are of American origin. According to the fair use doctrine, codified in Section 107 of the 1976 Copyright Act, the fair use of a copyright work is not an infringement if some requirements are met. There are four criteria to be assessed: 1. Purpose and character of the use of the work (including assessment on the commercial purpose). 2. Nature of the copyrighted work. 3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole. 4. The effect of the use upon the potential market for a value of the copyrighted work. For the official position of the US Copyright Office, see *Copyright and Artificial Intelligence, Part 3, Generative AI training*, May 2025.

[96] See Government of Canada, *What we heard report - Consultation on Copyright in the Age of Generative Artificial Intelligence*, 2025.

[97] *AI, copyright and the creative industries*, op. cit., p. 12.

[98] Zhang J. and Li X., *Fair Use in Training AI Models: A Review and Prospect of the Relevant Legal Development in China*, *Cambridge University Press*, 24 April 2026.

that have their place of establishment or are located within the Union and to providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union.[100]

The CDSMD does not provide specific rules in this regard.[101]

[99] According to Recital 106 AIA: "Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place. This is necessary to ensure a level playing field among providers of general-purpose AI models where no provider should be able to gain a competitive advantage in the Union market by applying lower copyright standards than those provided in the Union."

[100] Article 2 AIA.

[101] Zapala I., *Territorial scope of the authorization requirement and liability regime under Directive (EU) 2019/790 on copyright and related rights in the digital single market*, *Journal of Intellectual Property Law & Practice*, Volume 16, Issue 8, pp. 827-849, August 2021.

3. Copyright in AI training and AI prompting

3.1. AI training activities

Although the TDM exception under the European regime is frequently considered to cover the training of AI systems, there are diverging voices raising some concerns about the actual coverage of the whole process of training through Articles 3 and 4 CDSMD.

While the framework appears solid overall for the audiovisual industry,[102] academic doctrine suggests that AI training goes beyond TDM, being the latter part of the former, but not the other way round, and warns that no legal instrument has yet considered AI training and TDM as equivalent.[103] Also, doctrine warns that TDM tackles collection, processing and analysis of information but cannot cover important parts of AI training like those related to content.[104] From an institutional point of view, the European Union Intellectual Property Office (EUIPO) offers a more nuanced position balancing the different stances.[105] Among certain doubts concerning the real effects of the TDM exception, the "Voss Report" has also recently warned that the opt-out regime "may not cover all relevant acts of text and data mining". Some important rulings are being issued in Europe at national level: the most relevant cases at national level include that of the German collecting society GEMA v. OpenAI, where the Munich Regional Court ruled that ChatGPT had reproduced protected lyrics;[106] in other cases, such as Getty Images v Stability AI in the UK, some guidance was offered by the UK High Court but relevant questions have not yet been addressed.[107]

---

[102] See, for example, see the ACT website, *AI Principles for Audiovisual Media Services*, Brussels, 16 June 2026: "The EU's copyright and AI framework—particularly the CDSM Directive and the AI Act—sets out rules governing the use of protected works for AI training. Significant uncertainties around implementation remain, but the framework is robust enough to adapt to technological changes and provides a solid basis for assessing infringements at both the input and output stages of AI systems".

[103] See Rosati, E., *The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law*, op. cit. Rosati underlines that in TDM "the main acts under UK and EU law are those of extraction (if the data is incorporated in a database) and reproduction", not others.

[104] See Dornis, T., *The Training of Generative AI is not Text and Data Mining*, European Intellectual Property Review (E.I.P.R.), 19 October 2024.

[105] See *The development of Generative Artificial Intelligence from a Copyright Perspective*, op. cit., p. 67.

[106] For example, with respect to AI training, *GEMA v. OpenAI*, 42 O 14139/24, Landgericht München I.

[107] Stokes G. et al., *Getty Images v. Stability AI: the UK High Court decision offers guidance, but critical questions on AI and copyright infringement remain*, *DLA Piper*, 5 November 2025.

GenAI training can adopt several forms and subprocesses as it is a complex technical process. The latest GenAI models are multimodal as they can work on a variety of content types.[108] Typically, these systems build on all sorts of datasets, including videos, text or images that stem from private and public institutions or individuals and synthetic data. One of the popular ways of accessing this data is through web scraping and crawling.[109] Acquired data is then reprocessed to ensure its quality, including accuracy, consistency, relevance and other important aspects. Finally, training includes fine tuning before the generation of an output. As a consequence, in terms of copyright, looking at the different stages of AI training, the effects of copyright can be different, as the activities that fall within the scope of each of them are also different. According to academics, there are five relevant stages in AI training that entail different technical processes:

Figure 3. The technical process of training AI systems.

1. **Ingestion and preprocessing**, where content is collected.
2. **Training loop**, where the AI systems generate cleaner data.
3. **Model artefact**, where representations of the data obtained are generated but cannot be viewed.
4. **Generation and retrieval**, where user prompts are tokenised.
5. **Logs and caches**, where information is selected and partly cached for further training.

Source: Copyright in Formaldehyde: How GEMA v OpenAI Freezes Doctrine and Chills AI – Part 1 [110]

3.2. AI Prompting

Although prompting is not defined in legal instruments, it can be defined as the process of giving specific instructions or information to an AI system to obtain a particular result,

[108] Marcelin T. and Cassetti F., *AI and copyright: The training of general-purpose AI*, European Parliamentary Research Service, April 2025.

[109] As explained by EUIPO: "Crawling typically involves the systematic browsing of publicly accessible web pages to index metadata and content for search engines. This process is generally considered less intrusive and primarily serves discoverability functions. Scraping, by contrast, involves the extraction of specific data or content, often targeting more granular information and reproducing portions of works in a way that is more likely to raise copyright concerns". See *The development of Generative Artificial Intelligence from a Copyright Perspective*, op. cit., p. 139.

[110] See Frosio, G., *Copyright in Formaldehyde: How GEMA v OpenAI Freezes Doctrine and Chills AI – Part 1*, *Kluwer Copyright Blog*, 10 December 2025. See also, for a complete overview of the training process, the Generative AI tech stack from data collection to end-user interaction in the Council of Europe's *Guidance note on the implications of generative artificial intelligence for freedom of expression*, pp. 44-45.

taking into account that the clarity and precision of instructions or prompts will positively influence the quality of output.[111]

Prompts are the driver for obtaining an output in conversational chatbots. Apart from the debate around output copyrightability (for which prompts emerge, in GenAI and chatbot-like AI, as an indispensable tool, and which will be examined in the second part of this report), a second concern is the protectability of prompts as such, which is subject both to general copyright rules (which do not protect mere ideas) and possibly other areas of law like trade secrets.[112]

In the USA, the Copyright Office has concluded that "prompts alone do not provide sufficient human control to make users of an AI system the authors of the output. Prompts essentially function as instructions that convey unprotectible ideas. While highly detailed prompts could contain the user's desired expressive elements, at present they do not control how the AI system processes them in generating the output". In this vein, the US Copyright Office has warned that "the gaps between prompts and resulting outputs demonstrate that the user lacks control over the conversion of their ideas into fixed expression, and the system is largely responsible for determining the expressive elements in the output".[113]

[111] See the KU Leuven website, *Guiding AI with prompting* (last updated on 5 February 2026).
[112] For a reflection on trade secrets and AI, see Hrdy, C. *Trade Secrecy Meets Generative AI*, *Chicago Kent Law Review. Nr. 100,* 3 February 2025.
[113] US Copyright office, *Copyright and Artificial Intelligence, Part 2, Copyrightability*, January 2025.

4. Platforms and users: the terms of service

The law regulates (or partly regulates) the training of AI, and platforms usually include certain legal provisions in their terms of service that users accept when using them. In this connection, terms of service establish a private relationship between the user and the platform that governs several aspects like copyright and training. However, under European laws a platform may in any case incur liability under the individual legal regimes in force.

For example, in terms of European copyright the application of the TDM exception may result in the liability of an AI provider in violation of the opt-out regime. In other cases, liability may arise, for AI platforms, for the violation of other laws like the AIA and the DSA, as AI services are provided through digital platforms that may fall within its scope.[114]

Where terms of service exist, they must be fully considered but the liability of the provider cannot be excluded just because the terms of service confer responsibility upon the user.[115]

Some of the most popular GenAI platforms do in fact include some references to the input and the further use that the platforms can make of prompts and other material fed by users. Insofar as this responds to an acceptance of the terms and conditions by the user, and therefore constitutes a private and enforceable relationship, copyright implications may exist over the input, and the user usually accepts that the content can be used for training purposes.

As an example, these are some of the most frequent cases observed across popular GenAI tools:

[114] *Regulation (EU) 2022/2065* of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (DSA). According to Article 2, the DSA applies to intermediary services offered to recipients of the service that have their place of establishment or are located in the Union, irrespective of where the providers of those intermediary services have their place of establishment.

[115] See Rosati, E., *The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law*, op cit.

Table 5. AI training rules in the terms of service of some of the most relevant AI platforms.

Platform	Terms and conditions (relevant mentions)
Adobe[116]	Your prompts and inputs to, and the results generated by, generative AI features in Adobe products may be reviewed through both automated (such as machine learning) and manual methods for abuse prevention and content filtering purposes. Using Adobe's generative AI features to create, upload, or share content that violates third-party copyright, trademark, privacy, publicity, or other rights is prohibited. This may include entering text prompts designed to generate copyrighted, trademarked, or otherwise infringing content, uploading an input or reference image that includes a third party's copyrighted content, generating text that plagiarizes third-party content, or using a third party's personal information in violation of their privacy or data protection rights.
ChatGPT[117]	You may provide input to the Services ("Input") and receive output from the Services based on the Input ("Output"). Opt out. If you do not want us to use your Content to train our models, you can opt out by following the instructions in *this article*. Please note that in some cases this may limit the ability of our Services to better address your specific use case.
Claude[118]	It's up to you to choose whether to allow your data to be used to improve new Claude models and you can change your choice anytime in your Privacy Settings.
Copilot[119]	We don't own Your Content, but we may use Your Content to operate Copilot and improve it. By using Copilot, you grant us permission to use Your Content, which means we can copy, distribute, transmit, publicly display, publicly perform, edit, translate, and reformat it, and we can give those same rights to others who work on our behalf. We get to decide whether to use Your Content, and we don't have to pay you, ask your permission, or tell you when we do. But that doesn't mean we can use it however we want. The Microsoft Privacy Statement explains how we use Your Content, and the privacy options in Copilot give you control over some of those uses.
Midjourney[120]	By using the Services, You grant to Midjourney, its affiliates, successors, and assigns a perpetual, worldwide, non-exclusive, sublicensable no-charge, royalty-free, irrevocable copyright license to reproduce, prepare derivative works of, publicly display, publicly perform, sublicense, and distribute the Content You input into the Services, as well as any Assets produced by You through the Service. This license survives termination of this Agreement by any party, for any reason.

[116] See the Adobe website, *Adobe Generative AI User Guidelines* (last updated on 15 May 2026).
[117] See the OpenAI website, *Terms of Use* (Published on 1 January 2026).
[118] See the Anthropic website, *Updates to Consumer Terms and Privacy Policy* (published on 28 August 2025).
[119] See the Microsoft website, *Microsoft Copilot Terms of Use* (effective as of 12 June 2026).
[120] See the Midjourney website, *Terms of Service* (effective as of 27 May 2026).

5. Main takeaways

AI development continues to grow rapidly, with GenAI becoming increasingly integrated into professional and private uses and Agentic AI growing at a fast pace. Its impact is particularly strong in creative sectors, such as audiovisual production, where AI can generate text, images, audio and video. This rapid evolution creates new opportunities but also raises complex copyright challenges, especially under the European copyright framework, that legislators are trying to address.

AI training is at the core of the copyright and AI interplay, because AI models are trained on vast datasets that may include copyrighted works and other subject matter. The use of copyrighted content without authorisation triggers concerns among creators, publishers and cultural industries, such as the audiovisual sector, where cognitive tasks are important. The focus, from a policy point of view, is currently in balancing access to data with the protection of rightsholders.

In Europe, specific copyright legislation dedicated to GenAI can be found in some provisions in the AIA and the subsequent application of the CDSMD. The AIA introduces transparency obligations, while the AIA refers to the CDSMD for copyright matters. Future court decisions and legislative reviews are expected to clarify remaining uncertainties.

In terms of AI training, text and data mining (TDM) rules are at the centre of the debate, as they provide a possible legal basis for certain AI training activities. However, questions remain about whether these exceptions fully cover the complexity of AI training processes, which entail multiple technical processes. The effectiveness of rights reservation mechanisms (like the European "opt-out" mechanism) is also under discussion, as addressed, for example, in the recent "Voss Report".

Prompting also raises questions regarding copyright, particularly concerning the protection of prompts and the use of copyrighted materials. Since prompts mainly provide instructions rather than determine the final expression, current approaches question whether users exercise sufficient creative control over their materials in AI training according to prompting rules found in the Terms of Service of popular AI platforms. The relationship between human input and AI-generated content remains an evolving legal issue.

A publication
of the European Audiovisual Observatory