

MASTER THESIS | MASTER'S THESIS

Titel | Title

Evolution of Legislation and Judicial Practice in Applying Copyright Law to AI
Development in the European Union and the United States: Progress, Challenges,
and Future perspectives

verfasst von | submitted by

Ksenia Gnedkova

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Laws (LL.M.)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 999 082

Universitätslehrgang lt. Studienblatt | Postgraduate
programme as it appears on the student record
sheet:

Europäisches und Internationales Wirtschaftsrecht (LL.M.)
[Berufsbegleitend Englisch]

Betreut von | Supervisor:

Mag. Dr. Sascha Salomonowitz

ABSTRACT

Diese Arbeit untersucht das Zusammenspiel zwischen dem Urheberrecht und der künstlichen Intelligenz in der Europäischen Union und den Vereinigten Staaten. Sie befasst sich mit fünf Kernfragen: den anwendbaren Rechtsrahmen, dem urheberrechtlichen Status von KI-generierten Werken, der Rechtmäßigkeit des Einsatzes urheberrechtlich geschützter Materialien für das Training von KI, den aktuellen Positionen öffentlicher Stellen, Gerichten und Interessengruppen sowie der Notwendigkeit einer internationalen Harmonisierung.

Die Untersuchung stützt sich auf eine dogmatische und rechtsvergleichende Methodik, bei der gesetzliche Regelungen, politische Dokumente und eine auslegende Analyse einschlägiger Gerichtsentscheidungen in beiden Rechtsordnungen berücksichtigt werden.

Die Analyse zeigt, dass die EU ein harmonisiertes und präventives Modell verfolgt, und zwar durch den Erlass des AI Acts sowie die Anwendung der Text- und Data-Mining-Ausnahmen der DSM-Richtlinie, während die Vereinigten Staaten derzeit die Möglichkeit der Anwendung der Fair-Use-Doktrin und administrativer Auslegung erörtern. Beide Rechtsordnungen bekräftigen das Prinzip der menschlichen Urheberschaft; dennoch bleiben ungelöste Fragen hinsichtlich von Werken mit minimalem menschlichem Beitrag und des Eigentums an nicht geschützten KI-generierten Ergebnissen bestehen. Interessengruppen auf beiden Seiten lehnen ein unlizenziertes KI-Training ab und fordern Transparenz sowie eine angemessene Vergütung. Die Ergebnisse legen nahe, dass eine Harmonisierung auf nationaler und internationaler Ebene erforderlich ist, um Rechtssicherheit zu gewährleisten, die Interessen von Rechteinhabern und Entwicklern auszubalancieren und eine nachhaltige Innovation weltweit zu fördern.

Table of contents

List of abbreviations	VI
I. Introduction	1
1.1. Subject matter of the Master Thesis	1
1.2. Scope of the study and research questions	1
1.3. Limitations of the Master Thesis	2
1.4. Research methodology	3
II. AI and copyright-protected works	4
2.1. Introduction	4
2.2. How AI uses copyright-protected works, and how AI generates content similar to protected works	4
2.3. Legal challenges to copyright arising from AI development	7
2.3.1. Copyright protection for AI-generated content	7
2.3.2. The use of copyrighted works in AI training	8
2.3.3. Other related problems	9
2.4. Conclusion	10
III. Legislative framework for AI and copyright in the EU and the US	12
3.1. Introduction	12
3.2. Overview of existing EU copyright legislation applicable to AI	12
3.2.1. The development of EU copyright legislation applicable to AI	12
3.2.2 The EU AI Act and its implications for copyright law	16
3.2.3. Conclusion	19
3.3. Overview of US copyright law relevant to AI	20
3.3.1. Introduction	20
3.3.2. Federal approach	20
3.3.3. State's approach	22
3.3.4. Legislation perspectives	24
3.3.5. The role of the U.S. Copyright Office and its influence on the application of copyright law	25
3.3.6. Conclusion	27
3.4. Comparative analysis: EU vs US legislative approaches to AI and copyright	27

IV. Challenges and Complex Issues	30
4.1. Copyrightability of the AI-generated content	30
4.1.1. Introduction	30
4.1.2. Copyrightability under the EU legal framework	30
4.1.3. Copyrightability of AI-generated output under the EU legal framework	34
4.1.4. Copyrightability of AI-generated outputs under the US legal framework	36
4.1.5. Observations and conclusions on copyrightability of AI-generated output	40
4.2. The use of copyrighted works for AI training	42
4.2.1. Can copyrighted works be used to train AI? Legal challenges under the EU copyright framework	42
4.2.1.1. Machine-readable format	42
4.2.1.2. AI Training is not equal to TDM	44
4.2.1.3. Opt-out mechanism problems	44
4.2.1.4. From problem to policy: legislative recognition of the issues	45
4.3. The use of copyrighted works in AI training under US law: the fair use debate	46
4.3.1. What is fair use?	46
4.3.2. Arguments for fair use	48
4.3.3. Arguments against fair use	49
4.3.4. The position of the U.S. Copyright Office	50
4.3.5. Observations	52
4.4. Significant court cases: analysis of recent landmark court decisions	52
4.5. Approaches taken by other market participants, such as collective rights management societies.	59
4.6. Conclusion: comparative reflections on copyright, AI-generated content, and training uses in the EU and the US	62
V. Harmonisation	65
VI. Conclusion and Key Findings	69

Table of legislation, cases, and sources	73
1) Table of legislation	73
2) Table of Cases	74
3) Other Official Documents	76
4) Research, Studies and Surveys	77
Bibliography	78

List of abbreviations

AI	Artificial intelligence
CRS	Congressional Research Service
CJEU	Court of Justice of the European Union
CMO	Collective Management Organisation
EU	European Union
EWC	European Writers' Council
GPAI	Generative pre-trained artificial intelligence
LLM	Large language model
ML	Machine learning
TDM	Text and data mining
TFEU	Treaty on the Functioning of the European Union
TRIPS	Agreement on Trade-Related Aspects of Intellectual Property Rights
US	United States

I. Introduction

1.1. Subject matter of the Master Thesis

The rapid development of artificial intelligence, and in particular generative AI, has created both unprecedented opportunities and significant challenges for copyright law. Human authorship has traditionally been the foundation of copyright protection; however, the emergence of AI-generated content raises new legal questions about the copyrightability of works created wholly or partly by AI, as well as the lawful use of copyright-protected materials in the training of AI models. These issues are of particular importance to the creative industries, where the unauthorised exploitation of works may negatively affect the livelihoods of authors, performers, and other rightsholders.

Artificial intelligence has only recently become widely used, and the range of legal questions associated with it has emerged over the past few years. Until now, there has been neither the need nor the possibility to assess whether existing national legislations were adequate to regulate AI-related uses or whether additional legal measures were required. This challenge extends to all areas of law affected by AI, but copyright law has proven to be among the most profoundly impacted. The close interconnection between copyright-protected works and AI technologies has triggered intense debates, litigation, and calls for reform, highlighting the need for legislative action or, at the very least, consistent judicial interpretation and case law.

This thesis examines the legislative and judicial approaches to regulating artificial intelligence through copyright law in the European Union and the United States. These jurisdictions have established themselves as global leaders in shaping legal practice and legislation, particularly in the face of technological change, and have been chosen for this analysis for that reason. The study reviews current copyright legislation, case law, and emerging legal trends, and provides a comparative analysis of the EU's civil law framework and the US common law tradition.

Finally, the thesis identifies unresolved legal issues and explores possible solutions, with particular emphasis on the positions of institutional and market stakeholders whose interests are directly affected by the intersection of copyright law and AI.

1.2. Scope of the study and research questions

This research examines the legal frameworks in the European Union and the United States that govern AI-generated works and the use of copyright-protected materials in training artificial intelligence models. It analyses how existing laws in both jurisdictions define the legal status of AI-generated outputs and address the use of protected works in AI training datasets. The scope of the study includes statutory law, soft law instruments (recommendations, codes of practice,

guidelines, declarations, and statements of professional associations), and policy documents in both the EU and the US.

The main research questions are as follows:

- Which sources of law in the US and EU are applicable to the intersection of AI and copyright?
- To what extent are works produced by AI eligible for copyright protection under current EU and US law?
- How do the copyright regimes in both jurisdictions regulate the use of protected works for AI training?
- What positions do institutional actors, collective management organisations, and market stakeholders take on these matters?
- What are the principal similarities and differences between the EU and US approaches, and is there a need to amend legislation, create a distinct legal status for AI-generated works, or pursue harmonisation at the national and international levels?

Special emphasis is placed on court decisions, industry opinions, and official guidance from governmental bodies to evaluate how legal interpretation is evolving and how efforts are being made to balance technological innovation with the protection of copyright holders' rights. Ultimately, the study aims to assess the degree of legal certainty achieved to date and to identify gaps that continue to challenge stakeholders in the creative industries and AI development.

1.3. Limitations of the Master Thesis

The scope of this thesis is limited to identifying applicable legislation and its interpretation for the research questions and two core problems at the intersection of copyright and artificial intelligence: the copyrightability of AI-generated works and the use of copyright-protected materials for training AI systems. While the broader intersection of copyright and AI also includes questions of liability for infringing outputs, the protection and enforcement of moral rights, and the treatment of derivative works created with AI assistance, among others, these aspects fall outside the scope of this research.

The focus on these issues is dictated by their direct practical relevance to stakeholders in the creative industries and by the fact that they have already generated substantial legislative, judicial, and policy responses in the European Union and the United States. Limiting the scope in this way makes it possible to provide a more in-depth comparative analysis while ensuring the research remains manageable within the constraints of time and available resources.

1.4. Research methodology

The thesis combines doctrinal research, comparative legal analysis, and interpretive analysis.

Doctrinal research focuses on identifying, interpreting, and analysing the relevant primary sources of law such as statutes, regulations, directives, and courts' decisions, as well as secondary sources such as academic literature, studies, articles, and commentaries. This allows for a systematic understanding of how copyright law currently addresses (or fails to address) issues related to AI-generated works and the use of protected works in AI training.

Comparative legal analysis is applied to examine the approaches taken by the European Union and the United States, jurisdictions representing civil law and common law traditions, respectively. This method helps identify similarities, differences, and potential areas for mutual influence or harmonisation, particularly regarding the copyrightability of AI-generated works and the regulation of training datasets.

The research applies interpretive case law analysis to examine how courts in the EU and the US have approached issues related to AI-generated works and the use of copyright-protected material for AI training. This method goes beyond the mere description of judicial decisions and seeks to identify underlying legal reasoning, trends, and potential implications for future copyright regulation.

II. AI and copyright-protected works

2.1. Introduction

This Chapter outlines the basic principles of how generative AI works and main copyright-related challenges raised as a result of its large-scale dissemination and public accessibility, based on recent legislative developments, academic discourse in the US and the EU, and the key legal questions currently under active consideration by institutions such as the European Commission, the U.S. Copyright Office, legislators, and courts in both jurisdictions. The purpose of this Chapter is to designate the key problem areas without going into the details of ongoing debates, scholarly opinions, or specific legal provisions that have already been adopted in the United States and the European Union. A detailed analysis of those developments will be presented in Chapter IV, following an overview of the current legal framework and its emerging trends in Chapter III. It is therefore necessary to first establish the significance of this research focus as a basis for examining the applicable legal framework and conducting a deeper analysis of the copyright challenges it raises.

2.2. How AI uses copyright-protected works, and how AI generates content similar to protected works

Artificial intelligence systems, particularly those based on deep learning and generative models, are increasingly capable of producing content that is similar to human-created literary and artistic works. These include text (poetry, prose, journalism), images and paintings, music compositions, audio recordings, and even video and animation. In many cases, the outputs generated by these models are so sophisticated that they are indistinguishable from works protected by copyright under both international and national laws.

At the technical level, generative AI systems such as large language models (e.g., GPT from OpenAI, Claude from Anthropic, Gemini from Google DeepMind), which are specialized in processing texts — generating, translating, analyzing, answering questions, and so on — and diffusion models (e.g., DALL·E, which is integrated into ChatGPT Plus, Stable Diffusion from Stability AI, Sora from OpenAI), which are specialized in generating images and videos, are trained on massive datasets containing existing human-created content. These datasets may include books, artworks, photographs, music files, and other creative works, some in the public domain, others under copyright protection.

As explored in the research conducted by the U.S. Copyright Office¹, the datasets for AI training are being collected from a wide range of sources, including publicly available internet content, licensed datasets, and, in some cases, even infringing sources such as pirate libraries². Developers may scrape websites, extract structured databases (such as Wikipedia), or obtain curated datasets like Common Crawl. In addition, many incorporate proprietary or licensed data acquired through agreements with publishers, platforms, or data brokers. Regardless of the origin, this data is usually processed by filtering, cleaning, and formatting to prepare it for model training³.

AI content generation typically involves probabilistic modeling. In language models, the system predicts the next word based on the preceding context and the data it was trained on. As Meta Platforms, Inc. (Meta) explained in its defence against a copyright infringement claim, the model functions ‘by extracting increasingly complex mathematical patterns from training data, enabling the network to output a prediction or decision based on the patterns derived’⁴. In image models, visual elements are generated using text prompts or internal patterns learned during training that capture the key features, structure, and style of the desired image⁵. The resulting content is often new in form but often stylistically or structurally derivative.

One of the most well-known early examples demonstrating how AI can generate visual artworks is the 2016 project ‘The Next Rembrandt’⁶. The project used machine learning algorithms trained on scans of Rembrandt’s paintings to generate an entirely new portrait in Rembrandt’s style. The resulting image was printed in 3D to replicate Rembrandt’s brushstrokes. The project is widely cited as the first high-profile demonstration of AI’s ability to synthesize original content that is stylistically indistinguishable from the work of a specific human artist.

In the domain of music, projects such as AIVA (Artificial Intelligence Virtual Artist) and OpenAI’s MuseNet present similar opportunities. AIVA, for example, is an AI composer trained

¹ United States Copyright Office, ‘Copyright and Artificial Intelligence: Part 3 – Generative AI Training’ (pre-publication version, May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>> accessed 1 August 2025.

² *ibid*, 13-21.

³ *ibid*.

⁴ *Kadrey v Meta Platforms Inc*, No 23-cv-03417-VC, Order Denying Plaintiffs’ Motion for Partial Summary Judgment and Granting Meta’s Cross-Motion for Partial Summary Judgment, (N.D. Cal. 2025), 8. <<https://law.justia.com/cases/federal/district-courts/california/candce/3:2023cv03417/415175/598/>> accessed 05 August 2025.

⁵ Microsoft, ‘Prompt Engineering Techniques’ Microsoft Learn – Azure AI Foundry, (2025) <<https://learn.microsoft.com/en-us/azure/ai-foundry/openai/concepts/prompt-engineering?tabs=chat>> accessed 10 July 2025.

⁶ Chris Baraniuk, ‘Computer paints “new Rembrandt” after old works analysis’ (2016) *BBC News*, <<https://www.bbc.com/news/technology-35977315>> accessed 5 May 2025.

on a dataset of classical music scores. It has been used to generate original symphonic pieces and film music in the style of Beethoven, Bach, or Mozart⁷.

Users often guide AI systems to get the result by providing prompts, setting parameters, or reviewing and editing AI-generated outputs. A prompt is an instruction, reference, description of the idea, or other input that the user gives to the AI model to get the desired output. As noted on the OpenAI developer platform ‘prompting is both an art and a science’⁸. Some prompts are one-time, short, and simple requests, while others could be detailed and may be revised several times to achieve the intended result. The science of creating sophisticated prompts to guide AI models to generate desired results is now known as ‘prompt engineering’. Prompts can be automatically refined by AI systems to improve output quality. As it was mentioned in the second part of the report published by the U.S. Copyright Office ‘Copyright and Artificial Intelligence: Part 2 – Copyrightability’

‘although AI technology continues to advance, uncertainty around how a particular prompt or other input will influence the output may be inherent in complex AI systems built on models with billions of parameters. Some observers describe AI as a ‘black box’, and even expert researchers are limited in their ability to understand or predict the behavior of specific models’⁹.

While most AI-generated content is synthesized through probabilistic modeling and not direct copying, there have been confirmed cases in which the output produced by AI closely mirrors or even nearly duplicates specific works included in the training dataset. A group of researchers conducted experiments on several LLMs, including Meta’s Llama and found that memorization of training data varies significantly between models and between books. One of the Llama models demonstrated the highest level of memorization, capable of reproducing long excerpts from copyrighted books such as Harry Potter and the Sorcerer’s Stone and works by Camus, Tolkien, and Martin¹⁰.

In December 2023, the New York Times Company filed a copyright infringement lawsuit in the Southern District of New York against OpenAI and Microsoft¹¹, claiming that their AI models had

⁷ Gaurav Krishnan, ‘Could the Future of Music Composition Be AI? Analysing the Implications of AIVA — An AI Composer’ (2022) *Medium* <https://medium.com/@gaurav_krishnan/could-the-future-of-music-composition-be-ai-analysing-the-implications-of-aiva-an-ai-composer-8e3762354860> accessed 5 May 2025.

⁸ OpenAI, ‘Prompting: Overview – OpenAI API’ *OpenAI Platform Documentation* <<https://platform.openai.com/docs/guides/prompting>> accessed 2 July 2025.

⁹ U.S. Copyright Office, ‘Copyright and Artificial Intelligence. Part 2: Copyrightability’, (2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>> accessed 10 August 2025.

¹⁰ A. Feder Cooper and others, ‘Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models’ (2024) <<https://arxiv.org/abs/2505.12546>> accessed 31 July 2025

¹¹ The New York Times Company v Microsoft Corporation and OpenAI Inc, No 1:23-cv-11195, (SDNY, December 2023).

reproduced the New York Times' content nearly verbatim, including full articles, without permission.

Thus, even though AI developers most likely did not intend for their models to reproduce others' works, exact reproductions do occur in practice, as it is difficult even for the developers themselves to fully understand how AI systems generate their outputs because of their 'black box' nature.

Thereby, the development, training, and functioning of AI systems are deeply intertwined with the use of copyright-protected works. AI technologies are also capable of generating outputs that fall within the same categories of works, such as literary texts, images, music, or audiovisual content, that are traditionally protected by international and national copyright laws. This raises a number of legal and ethical concerns that are currently the focus of widespread debate and will be further described in the following Chapter.

2.3. Legal challenges to copyright arising from AI development

The widespread development and deployment of AI, particularly generative models, have introduced a range of unresolved legal questions concerning copyright law. As AI technologies become increasingly capable of generating content typically protected under international and national copyright frameworks, they challenge the boundaries of existing legal doctrines and create uncertainty for creators, developers, users, rights holders, and other stakeholders. This section focuses on identifying the specific copyright issues that have emerged as a result of the development of generative artificial intelligence, affecting the creative industries and AI developers.

2.3.1. Copyright protection for AI-generated content

A central complicated legal issue is whether outputs generated by AI systems qualify for copyright protection under existing laws. Copyright law has long been based on the principle of human creative work, both in the EU and the US. The case law in both jurisdictions requires that a work reflect the author's own intellectual effort and creative choices. Both legal systems define the author as a human being, with limited exceptions in specific cases.¹² However, the development of generative AI challenges these rules, since most of such outputs are produced by algorithms without direct human authorship or minimal human involvement, thereby raising doubts about

¹² See, e.g., Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20 (Database Directive), art 4(1), which provides that the author of a database may be the legal person under whose name the database was made; or Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs [2009] OJ L111/16 (Computer Programs Directive), article 2(1), which provides that the 'author of a computer program shall be the natural person or group of natural persons who has created the program or, where the legislation of a Member State so permits, the legal person designated as the rightsholder by that legislation'.

their eligibility for copyright protection. This raises the question: can a work created with minimal or no human involvement qualify for copyright protection at all?

Some scholars argue that addressing this question may require a fundamental rethinking of copyright doctrine and legislation. For example, Mark Lemley, in his essay ‘How Generative AI Turns Copyright Law Upside Down’¹³, claims that existing legal concepts such as the idea-expression dichotomy and the substantial similarity test no longer function effectively in the context of AI-generated content. Since AI performs the expressive work which is traditionally rewarded by copyright, and a human input at the same time is often limited to prompts or ideas, the current legal framework struggles to determine what, if anything, is protectable. Lemley suggests that this shift may require rethinking not only how authorship and infringement are assessed, but also the very foundations of what copyright is meant to protect.

In addition, some AI-generated outputs are created without any human creative input, others involve minimal human contribution, and still others are created by a human who uses AI merely as a tool. These types of human involvement may lead to different legal regimes related to the output, which is why the issue is so complex and widely debated.

2.3.2. The use of copyrighted works in AI training

Another major legal question is whether AI developers may lawfully use copyright-protected materials for training purposes. As it was mentioned above, most generative AI systems are trained on huge datasets, including those that are scraped from the internet, which often include works protected by copyright. During training, these models ingest and analyze the structure, style, patterns of a wide range of creative content (books, music, code, artworks, etc.). While this process does not involve copying in the traditional sense of reproducing a work for further distribution, it is still the reproduction and storage of the work in the database and the computer memory.

The central legal question is whether such use constitutes copyright infringement. In jurisdictions like the US and the EU, this issue is further complicated by the applicability of exceptions, such as fair use in the US and text and data mining in the EU. An additional question arises as to how traditional defences like fair use or TDM exceptions apply to automated, large-scale volumes of copying and usage for training of the models. Courts and lawmakers are currently trying to determine whether the use of protected works as training material without a license or other

¹³ See, Mark A. Lemley, ‘How Generative AI Turns Copyright Upside Down’ (2024), 25, Science and Technology Law Review <https://law.stanford.edu/wp-content/uploads/2024/09/2024-09-30_How-Generative-AI-Turns-Copyright-Upside-Down.pdf> accessed 3 June 2025.

authorisation of the rightsholders can be justified under fair use, TDM, or fall within any other legal exemption. These issues will be addressed in detail in Chapter IV.

Another related concern is that the economic benefits derived from such training do not translate into transparent or effective compensation for rightsholders whose works are used in the process. At present, there is no clear answer as to whether creators should receive fair remuneration for the use of their works in AI training.

2.3.3. Other related problems

There are also several other issues related to the use of copyright-protected works in AI training that fall outside the scope of this thesis. Nevertheless, they must be acknowledged in order to highlight the broader context and the interconnected nature of these challenges, which should be taken into account by legislators and courts, especially when attempting to address the questions outlined in sections 2.3.1 and 2.3.2 above.

First, there is an ongoing discussion about whether AI-generated content constitutes a derivative work under copyright law when it is based directly or indirectly on protected works¹⁴. Given that AI models learn from extensive training materials, there is a risk (as was noted above) that they may reproduce or closely imitate pre-existing works. In some cases, outputs may include elements (e.g., specific phrases, melodies, or visual compositions) that resemble or replicate parts of copyrighted works used during training. Thus, such use may raise concerns about the unauthorized production of derivative works.

Moreover, if a user instructs the AI to generate content ‘in the style of’ a particular creator, courts may have to consider whether this could amount to infringement¹⁵. While artistic styles themselves are generally not protected under copyright law, such prompts may result in works that are substantially similar to original works by that creator, thereby raising potential legal concerns¹⁶.

Second, the potential for AI systems to infringe upon copyright raises further legal issues regarding liability. Questions remain as to who should be held responsible when an AI output infringes on a copyrighted work: the developer of the model, the user who generated the content, or possibly both?

¹⁴ See, e.g., Claudio Novelli and others ‘Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity’ (2024) 55 *Computer Law & Security Review* 106066, para 4.2; Nicola Lucchi, ‘Generative AI and Copyright: Training, Creation, Regulation’ (Study for the European Parliament, Policy Department for Citizens’ Rights and Constitutional Affairs, 2025) 107-108.

< [https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU\(2025\)774095_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf)>
accessed 12 July 2025.

¹⁵ Lucchi (n 14) 107.

¹⁶ *ibid.*

Third, there are ethical and moral rights questions. As it was specified in the Resolution on intellectual property rights for the development of artificial intelligence technologies ‘a human-centered approach to AI that is compliant with ethical principles and human rights is needed if the technology is to remain a tool that serves people and the common good’¹⁷. This is particularly relevant in discussions about digital replicas and deepfakes, where an artist's voice, image, or persona may be simulated by AI without consent, raising issues under personality rights frameworks.

Fourth, in the process of using copyright-protected works for AI training, moral rights may also be infringed, particularly the right of integrity¹⁸. This right allows authors and performers to object to any use of their work that they perceive as a distortion or other derogatory treatment. Even when the work is not recognisable in the AI-generated output, and even if the economic rights have expired or been transferred, the use of the work as input data alone may still be problematic. For example, authors or their heirs might consider the transformation of their work by an AI system, without their consent, as disrespectful or contrary to the original meaning or intention of the work¹⁹.

Another actively discussed issue both in the United States and the European Union is the potential market impact of AI-generated outputs that may substitute for copyright-protected works. In the US, this concern is debated in the context of the fair use doctrine. In the EU, the debate tends to focus more broadly on the economic consequences and structural effects on creative industries and the livelihoods of entire groups of authors and performers. The concern is that the large-scale deployment of AI-generated content could replace human creators, reshape market dynamics, and destroy or significantly change existing models of creative production and remuneration²⁰.

2.4. Conclusion

As this Chapter has shown, the copyright challenges associated with AI are diverse and evolving. They include fundamental questions about the scope of protection, concept of the authorship, fair use, as well as issues of infringement, derivative works, and ethical concerns. While many of these problems are interrelated, this thesis focuses primarily on the issues of authorship and the use of

¹⁷ European Parliament, Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (2020/2015(INI)) https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html accessed 6 August 2025.

¹⁸ See, European Commission: Directorate-General for Communications Networks, Content and Technology, Crowell&Moring, IMC University of Applied Sciences Krems, Philippe Rixhon Associates, Technopolis Group and UCLouvain, ‘Study on copyright and new technologies – Copyright data management and artificial intelligence’, Publications Office of the European Union, 2022, 190 <<https://data.europa.eu/doi/10.2759/570559>> accessed 03 August 2025.

¹⁹ *ibid.*

²⁰ See, e.g., Lucchi (n 14), 14; European Commission, *Study on copyright and new technologies* (2022) (n 18) 149.

copyright-protected materials in training, as they form the foundation for addressing all other legal uncertainties in this field.

III. Legislative framework for AI and copyright in the EU and the US

3.1. Introduction

The legal debate over whether computers can be considered authors or inventors has been ongoing for at least thirty years²¹. Nevertheless, the emergence of works created with the help of computers did not raise significant legal concerns, as the computer was still primarily used as a tool, and the human creative contribution remained predominant. This is confirmed by the fact that these early discussions largely remained within academic and theoretical circles, without leading to legislative changes or widespread judicial debates.

The introduction of publicly accessible generative artificial intelligence changed everything. As discussed above the types of works generated by AI are listed in the Berne Convention²² as objects of copyright, but human involvement in their creation, in many cases, has been reduced to a minimum.

However, as noted above, since these works are not always created by a human but often merely involve their participation, questions arise as to whether the existing copyright framework applies to such works, or whether adjustments to the current legislation or the creation of new legal norms are required to adequately address them.

This Chapter examines which legislation currently applies in the European Union and the United States to works created by AI and what amendments, if any, have been introduced in response to this emerging technology.

3.2. Overview of existing EU copyright legislation applicable to AI

3.2.1. The development of EU copyright legislation applicable to AI

To date, European Union copyright law has not changed significantly in response to the development of artificial intelligence. In determining whether a work is protected by copyright, we apply the same legal acts and established case law as we do for works created exclusively by humans. Before the adoption of the Digital Single Market Directive²³ in 2019, there were only

²¹ See, e.g., Mark A. Lemley and Bryan Casey, 'Fair Learning' (2021) 99 *Texas Law Review*, 743; Pamela Samuelson, 'Allocating Ownership Rights in Computer-Generated Works' (1985) 47 *University of Pittsburgh Law Review* 1185.

²² Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, as amended 28 September 1979) 828 UNTS 221.

²³ 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 [2019] OJ 2019 L 130, 92.

European Commission Communications²⁴, Expert Groups and Committee's opinions²⁵ on the European level.

These documents mostly outlined the European Commission's strategy for fostering the development and responsible use of artificial intelligence in the EU, focusing on investment, ethical standards, and socio-economic adaptations. However, they barely mentioned copyright and other intellectual property issues related to AI development.

For example, the Opinion of the European Economic and Social Committee identifies eleven 'areas where AI has challenges: ethics; safety; privacy; transparency and accountability; work; education and skills; (in)equality and inclusiveness; law and regulations; governance and democracy; warfare; and superintelligence'²⁶. Notably, the intellectual property is not even mentioned among these challenges.

In the Communication on 'Artificial Intelligence for Europe'²⁷ it is only mentioned that the 'reflection will be needed on interactions between AI and intellectual property rights, from the perspective of both intellectual property offices and users, with a view to fostering innovation and legal certainty in a balanced way'. However, it is crucial to note that it was also stated in this Communication that 'The European Union (EU) should have a coordinated approach to make the most of the opportunities offered by AI and to address the new challenges that it brings'²⁸, which indicates that the European Union considers it necessary to harmonise AI-related legislation at the EU level.

In 2019 the 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 ('DSM Directive') was adopted. This Directive introduced the first amendments to EU legislation that influenced the development of AI in the European Union, although at the time of its adoption, generative AI tools, including models like ChatGPT, had not yet appeared on the market. These amendments are set out in Articles 3 and 4 of the DSM Directive, which introduced exceptions to the exclusive right of reproduction for the purposes of text and data mining. Article 3 provides for a mandatory exception

²⁴ See, e.g., European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe, COM (2018) 237 final; European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Coordinated plan on Artificial Intelligence COM (2018) 795 final.

²⁵ See, European Commission, AI High Level Expert Group, <<https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>> accessed 15 July 2025

²⁶ European Economic and Social Committee, 'Opinion on Artificial Intelligence — The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society' [2017] OJ C 288/01, para 1.5.

²⁷ European Commission, 'Artificial Intelligence for Europe' (n 24) 14.

²⁸ *ibid*, 2.

applicable to research organisations and cultural heritage institutions, allowing them to make reproductions and extractions of protected works and other subject matter to which they have lawful access, to carry out text and data mining strictly for scientific research. Article 4 establishes an additional exception that applies to all users, including commercial actors, provided they have lawful access to the content and allow ‘reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining’. However, rightsholders may opt out of this exception by explicitly reserving their rights in an appropriate manner, such as through machine-readable means in the case of publicly available online content.

As discussed above, at the time when DSM Directive was being drafted and adopted — between 2016 and 2019 — general-purpose artificial intelligence systems, such as ChatGPT and other large-scale generative models, were not yet publicly available or widely discussed in legal and policy frameworks. However, as it will be shown further in this thesis, with the emergence of such models, the text and data mining exceptions set out in Articles 3 and 4 have become central to the legal framework governing the use of publicly accessible content for the development of AI in the European Union.

In February 2020 the European Commission published the ‘White Paper on Artificial Intelligence - A European approach to excellence and trust’²⁹. In this document, the European Commission pointed out that ‘while the EU legislation remains in principle fully applicable irrespective of the involvement of AI, it is important to assess whether it can be enforced adequately to address the risks that AI systems create, or whether adjustments are needed to specific legal instruments’. According to the document, ‘the main risks related to the use of AI concern the application of rules designed to protect fundamental rights (including personal data and privacy protection and non-discrimination), as well as safety and liability-related issues’. Accordingly, a process was initiated to assess and adjust the legal framework to ensure it is fit for addressing the challenges and risks arising from the use of artificial intelligence. As is evident, intellectual property issues were once again excluded from the key areas of concern in the regulation of artificial intelligence.

In November 2020, the European Commission published an IP Action Plan³⁰, which announced the launch of a study on copyright and new technologies, specifically focusing on copyright data management and artificial intelligence. The same year, the European Commission published a report ‘Trends and Developments in Artificial Intelligence — Challenges to the Intellectual

²⁹ European Commission, ‘White Paper on Artificial Intelligence – A European approach to excellence and trust’, COM (2020) 65, final.

³⁰ European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Making the most of the EU’s innovative potential: An intellectual property action plan to support the EU’s recovery and resilience COM (2020) 760 final.

Property Rights Framework³¹ (2020), which provides detailed analyses on the copyrightability of AI-generated content and the results of which will be analysed in Chapter IV.

These initiatives demonstrated a growing institutional concern about the implications of AI for copyright law and the need to assess whether the existing legal framework remains fit for purpose in light of technological developments.

This was also reflected in the Resolution on artificial intelligence in education, culture, and the audiovisual sector issued by the European Parliament in May 2021³². The European Parliament declared that the existing intellectual property framework is generally suitable for supporting AI development, but requires targeted adaptations to respond to new challenges. However, the Parliament also expressed concern over the legal uncertainty surrounding authorship and ownership of AI-generated works and encouraged the European Commission to consider a common, evidence-based, and technologically neutral approach to address these issues.

In 2021, the European Commission published a ‘Proposal for a Regulation laying down harmonised rules on artificial intelligence’³³, which was followed by the active development of this Regulation, which was adopted and formally approved by the European Council on 24 May 2024.

Thus, within a relatively short period, the European Union adopted a specialised legislative act governing artificial intelligence. This act contains several provisions relevant to copyright, which will be examined in the sections below, but they relate to a specific area of exceptions to the exclusive right of reproduction for the purposes of text and data mining.

As for other matters not addressed by the AI Act, they continue to be governed by the copyright legislation adopted before the AI Act entered into force, including both general rules on copyright and those concerning specific subject matters, in particular the DSM Directive, the Term Directive, the Database Directive, the Computer Programs Directive, the InfoSoc Directive³⁴.

³¹ European Commission, ‘Trends and Developments in Artificial Intelligence — Challenges to the Intellectual Property Rights Framework’ (2020), <<https://digital-strategy.ec.europa.eu/en/library/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-framework>>, accessed 14 June 2025.

³² European Parliament, Resolution of 19 May 2021 on artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI) <https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_EN.html> accessed 14 June 2025.

³³ 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.

³⁴ 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 [2001], OJ L 167/10 (InfoSoc Directive); Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights [2006] OJ L 372/12 (Term Directive); Computer Programs Directive (n 12), Database Directive (n 12).

3.2.2. The EU AI Act and its implications for copyright law

The Artificial Intelligence Act³⁵ (AI Act) is the first legislative instrument of its kind worldwide, adopted by the European Union to comprehensively regulate artificial intelligence. With this act, the EU has taken a pioneering step towards establishing a legal framework that seeks to balance the protection of rights of developers, users, and other stakeholders with the need to support innovation and technological progress. Notably, the AI Act introduces, for the first time at the legislative level, specific provisions concerning the use of copyright-protected content for the training of AI systems. The following analysis focuses on the specific provisions of the AI Act relating to copyright-protected content and assesses their impact on the existing legal framework.

As emphasised in the recital 105 of the AI Act

‘general-purpose AI models, in particular large generative AI models, capable of generating text, images, and other content, present unique innovation opportunities but also challenges to artists, authors, and other creators and the way their creative content is created, distributed, used and consumed’.

Subsequently, the Regulation sets out a general principle that the use of copyright-protected material requires authorisation from the rightholder, unless an applicable exception or limitation under EU law applies. These exceptions are referenced in Article 53(1)(c), which establishes an obligation for providers of general-purpose AI models to implement a policy ensuring compliance with ‘Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790’.

Thus, the AI Act does not introduce a new substantive rule regarding the training of AI systems but extends the applicability of Article 4(3) of the DSM Directive to using copyright-protected content for AI training purposes. However, the legislator further developed this provision within the AI Act by imposing additional obligations on AI developers in connection with the application of this article.

These additional obligations, as set out in recitals from 105 to 108 and Article 53(1)(c) of the AI Act, significantly expand the responsibilities of providers of general-purpose AI models concerning the use of copyright-protected content. First, providers are required to adopt and implement a comprehensive internal policy for ensuring compliance with EU copyright and related

³⁵ 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) [2024] OJ L1689/1.

rights legislation. This policy must specifically address the obligation to identify and respect all reservations of rights expressed by rightsholders under Article 4(3) of Directive (EU) 2019/790.

Furthermore, the AI Act introduces a transparency obligation, whereby providers must draw up and make publicly available a sufficiently detailed summary of the content and data used to train their AI models. As clarified in recital 108, this summary does not need to contain exhaustive technical details, but must be comprehensive enough to enable copyright holders and other interested parties to protect and enforce their rights under EU law. The summary should, for instance, identify major data collections, datasets, or sources (whether public or private) and provide a narrative explanation of the nature and origin of the training material. These documentation and disclosure obligations are further detailed in Annex IX of the AI Act, which specifies the structure and content of the technical documentation required for general-purpose AI models. Among other things, the Annex requires the inclusion of a description of the model's training approach, information on the data used for training, testing, and validation, the sources of training data, and the description of the data selection and filtering process.

For consistent implementation and enforcement of the AI Act throughout the European Union, the European AI Office was established by the European Commission in early 2024 pursuant to Article 56 of the AI Act. In order to ensure the effective fulfillment of Article 53, the AI Office developed a 'Template for the Public Summary of Training Content for General-Purpose AI Models'³⁶, 'Code of Practice for General-Purpose AI Models'³⁷, and 'Guidelines on the scope of the obligations for general-purpose AI models established by AI Act'³⁸. The Code of Practice contains the Copyrights Section, which

'aims to contribute to the proper application of the obligation of providers of general purpose AI models placed on the Union market to put in place a policy to comply with Union law on copyright and related rights pursuant to Article 53(1), point (c) AI Act'.

The Code of Practice for General-Purpose AI Models is a non-binding, voluntary instrument. Providers may choose to sign on to the Code in order to demonstrate compliance with Article 53 obligations, particularly in relation to transparency and copyright. While signing the Code is not mandatory, it is anticipated by the AI Office as a form of industry self-regulation.

³⁶ European Commission, 'Template for the Public Summary of Training Content for General-Purpose AI Models' (2025) <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>> accessed 3 August 2025.

³⁷ European Commission, 'Code of Practice for General-Purpose AI Models' (2025) <<https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>> accessed 3 August 2025.

³⁸ European Commission, 'Guidelines on the scope of the obligations for general-purpose AI Models established by AI Act' (2025) <<https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>> accessed 3 August 2025.

Thus, the AI Act gives practical force to Article 4(3) of the DSM Directive by incorporating it into a clearer and more enforceable regulatory legislation. Providers are held directly responsible not only for respecting opt-outs, but also for explaining how copyrighted content may have been used during the development of their models, which corresponds to the request for transparency. It will ensure that creators and rights holders are able to verify whether their works were used for training purposes. The AI Act also emphasised that providers must use state-of-the-art technological means to detect opt-outs, ensuring that such restrictions are identified even when they are not easily recognisable through standard automated processes. This requirement aims to prevent developers from ignoring potential limitations simply because they are not immediately visible or explicitly stated. However, as it will be discussed in Chapter IV, case law may influence the interpretation and practical application of these provisions.

The AI Act also establishes that the obligation to comply with copyright and related rights legislation applies to all providers of general-purpose AI models who place their systems on the EU market, regardless of where the training of the model takes place. This provision aims to prevent providers from bypassing EU standards by outsourcing training to jurisdictions with lower copyright protection standards, and to ensure fair competition within the internal market.

Together, the AI Act's provisions reflect the EU's intention to integrate copyright safeguards directly into the regulatory treatment of AI models, particularly those trained on large-scale datasets, and to ensure that rightsholders retain effective control over the reuse of their works in the context of AI development.

While the adoption of the AI Act has raised concerns that the European Union may be introducing a complex and burdensome regulatory framework, potentially discouraging some developers from entering or remaining in the EU market, such fears have so far not resulted in widespread negative reactions. This may be due to the fact that the AI Act has only recently come into force, and how its rules will be applied in practice remains unclear. Nevertheless, following the formal adoption of the Regulation, major actors such as OpenAI have publicly expressed their intention to comply with the new rules³⁹, confirming a broader willingness within the industry to adapt to the emerging European standards.

³⁹ OpenAI, 'Complying with the EU AI Act's Prohibited Practices Requirements' <<https://help.openai.com/en/articles/10658393-complying-with-the-eu-ai-acts-prohibited-practices-requirements>> accessed 3 August 2025.

3.2.3. Conclusion

As discussed above, the European Union has so far chosen not to adopt a new copyright framework in response to the rise of artificial intelligence. Instead, existing legislation continues to serve as the primary legal basis governing the use of protected works in the context of AI.

The IP Action Plan⁴⁰ published by the European Commission acknowledges the growing importance of AI and its implications for intellectual property law. While the plan confirms that the current EU IP framework is broadly suitable for dealing with AI-assisted creations, it also points out that ‘there is still room for improvement and further harmonisation’. The Commission recommends a careful, evidence-based approach, focusing on stakeholder consultations and impact assessments before considering any legislative changes.

Similarly, the 2020 White Paper on AI⁴¹ supports the idea that regulation in this field must remain flexible and should only address clearly identified problems for which practical solutions exist.

Nowadays, the European authorities acknowledge the complexity of the issue and continue to monitor the situation by conducting new studies⁴² and surveys⁴³ of Member States in order to identify emerging problems. In particular, a 2024 survey⁴⁴ conducted by the Council of the European Union revealed that Member States are primarily concerned with the applicability of the TDM exception to AI training. Member States also suggest that, before introducing new transparency obligations related to AI-generated content, it is important to first analyse how the provisions of the AI Act will be implemented in practice. The majority of Member States consider that introducing a legislative instrument on the copyright aspects of AI at this stage is not necessary. Instead, the existing legal framework should be implemented, applied, and carefully monitored. Most Member States agree that international discussion on the topic of AI and copyright is needed and that the EU should take an active role in it. Thus, while legislative change is not required, the emphasis on ‘at this stage’ suggests a shared understanding among both EU institutions and Member States that changes to the law may still be needed in the future.

⁴⁰ IP Action Plan (n 30).

⁴¹ White Paper on Artificial Intelligence (n 29).

⁴² Lucchi, (n 14).

⁴³ Council of the European Union, ‘Revised summary of Member States contributions on the policy Questionnaire on the relationship between generative artificial intelligence and copyright and related rights Prepared by the Hungarian Presidency’ 16710/1/24 REV 1 (2024) <<https://data.consilium.europa.eu/doc/document/ST-16710-2024-REV-1/en/pdf>> accessed 7 August 2025.

⁴⁴ *ibid.*

Thus, given the scale and impact of the issue, which could transform both creative markets and technologies and affect society as a whole, the current approach is one of close observation and cautious evaluation.

3.3. Overview of US copyright law relevant to AI

3.3.1. Introduction

The United States has not yet adopted a comprehensive federal legal framework specifically addressing the intersection of artificial intelligence and copyright. Instead, legal developments are represented by a combination of federal agency initiatives, executive orders, state-level legislation, and the evolving role of the U.S. Copyright Office. This section provides an overview of key legislative and regulatory developments in the United States related to the interaction of copyright and AI, with a focus on both federal and state-level approaches, as well as the institutional role of the U.S. Copyright Office in shaping copyright policy in the context of copyright and AI. It also examines current legislative perspectives and the growing importance of judicial interpretation in addressing unresolved legal questions.

3.3.2. Federal approach

Initial discussions around the need for AI-specific regulation in the US began to emerge around 2016–2017. In 2016, the Obama Administration published a report ‘Preparing for the Future of Artificial Intelligence’⁴⁵ and ‘The National Artificial Intelligence Research and Development Strategic Plan’⁴⁶. These reports primarily emphasised the strategic importance and potential of AI technologies and their development in different industries. However, they also acknowledged several emerging risks, such as algorithmic bias, lack of transparency and explainability in AI decision-making, safety and reliability failures, and the potential for widespread job displacement due to automation. As in the European context, intellectual property infringement and copyright concerns were not yet part of the discussion at that time.

With regard to legislative regulation, a risk-based approach was recommended in the aforementioned documents in order to assess the potential of artificial intelligence to either reduce or increase certain risks in specific products or services. Based on this assessment, it should then

⁴⁵ Executive Office of the President, National Science and Technology Council, Committee on Technology ‘Preparing for the Future of Artificial Intelligence’ (2016) <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> accessed 7 July 2025.

⁴⁶ National Science and Technology Council, Select Committee on Artificial Intelligence ‘National Artificial Intelligence Research and Development Strategic Plan’ (2023 Update) <<https://www.nitrd.gov/pubs/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>> accessed 7 July 2025.

be determined whether the existing regulations already adequately address the risk, or whether they need to be adapted to AI⁴⁷.

At the federal level, the United States has only introduced AI-related regulatory updates within specific areas of law governing individual sectors. For example, the National Defense Authorization Act for Fiscal Year 2019⁴⁸ included provisions establishing an independent commission to review advances in artificial intelligence, related machine learning developments, and associated technologies. The National AI Initiative Act established a coordinated federal strategy for AI research and governance⁴⁹.

Several US federal agencies have taken sector-specific actions to address the risks and implications of AI technologies within their regulatory scope. For example, in the labor sector, the Equal Employment Opportunity Commission (EEOC) issued guidance under the Americans with Disabilities Act (ADA)⁵⁰, focusing on how algorithmic hiring tools may unintentionally discriminate against individuals with disabilities. The Federal Trade Commission (FTC) has applied consumer protection laws to deceptive or manipulative uses of AI⁵¹.

However, despite these efforts in sector-specific domains, no significant statutory amendments have yet been made in the US on the federal level, and all the concerns, including IP-related, continue to be addressed primarily through administrative guidance and ongoing public consultations, rather than through binding legislative reform.

At the presidential level, AI policy direction in the US is set by the current administration. Under President Biden, Executive Order 14110 ‘Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence’⁵² established a broad, risk-aware strategy requiring agencies to implement safety testing, protect workers and consumers, develop standards, and monitor bias and privacy. The Order also directed the U.S. Copyright Office and other agencies to study the impact of AI on intellectual property rights.

⁴⁷ Preparing for the Future of Artificial Intelligence (n 45), 17.

⁴⁸ H.R.5515 - 115th Congress (2017-2018): John S. McCain National Defense Authorization Act for Fiscal Year 2019, H.R.5515, 115th Cong. (2018), <<https://www.congress.gov/bill/115th-congress/house-bill/5515>> accessed 8 July 2025.

⁴⁹ H.R.6216 - 116th Congress (2019-2020): National Artificial Intelligence Initiative Act of 2020, H.R.6216, 116th Cong. (2020), <<https://www.congress.gov/bill/116th-congress/house-bill/6216/text>> accessed 8 July 2025.

⁵⁰ U.S. Department of Justice, ‘Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring’ <<https://www.ada.gov/assets/pdfs/ai-guidance.pdf>> accessed 8 July 2025.

⁵¹ Federal Trade Commission, ‘FTC Announces Crackdown on Deceptive AI Claims and Schemes’ (2024) <<https://www.ftc.gov/news-events/press-releases/2024/07/ftc-announces-crackdown-on-deceptive-ai-claims-and-schemes>>, accessed 08 July 2025.

⁵² Executive Office of the President, Executive Order No 14110, ‘Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence’, 88 FR 75191, (2023) <<https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>> accessed 09 July 2025.

In sharp contrast, President Trump's January 2025 'Removing Barriers to American Leadership in AI' Executive Order (EO 14179)⁵³ canceled Biden's order and directed agencies to revise or eliminate any measures seen as obstacles to US AI leadership or innovation. The new Order directs federal departments and agencies to review and, where necessary, revise or revoke any policies, regulations, or actions introduced under the Biden administration that are deemed inconsistent with enhancing America's global leadership in artificial intelligence⁵⁴. This shift highlights the different approaches: Biden focused on reducing risks and protecting the public, while Trump emphasised cutting restrictions to speed up AI development. As will be further discussed in this thesis, President Trump's approach to AI governance may have implications for copyright protection, especially when compared to the policy direction that had begun to emerge before his return to the office.

3.3.3. State's approach

While the federal government is initially cautious about introducing sector-wide AI legislation, US states have taken a more proactive approach. In recent years, an increasing number of US states have proposed legislation addressing artificial intelligence. The most interesting for our topic are the following legislative acts.

In April 2025, the state of Arkansas enacted Act 927 (HB 1876)⁵⁵, introducing one of the first comprehensive state-level rules on ownership in the context of generative AI. According to this Act the person who provides the input or directive to the generative AI tool to generate a content 'shall be the owner of the generated content, provided that the content does not infringe on existing copyrights or intellectual property rights'⁵⁶ and the person who provides data or input to train a generative AI model 'shall be the owner of the resulting trained model, provided that the training data is lawfully acquired and person has not transferred ownership rights through a contract or agreement'⁵⁷. The law makes an exception from these rules for the content that is generated as part of an employee's job duties and under the direction and control of an employer; in this case,

⁵³ Executive Office of the President, Executive Order No 14179, 'Removing Barriers to American Leadership in Artificial Intelligence', 90 FR 8741 (2025) <<https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>> accessed 7 July 2025.

⁵⁴ White & Case LLP, 'AI Watch: Global Regulatory Tracker – United States', (2025) *White & Case* <<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states>> accessed 29 July 2025.

⁵⁵ Arkansas, House Bill 1876, 95th General Assembly, An Act Regarding the Ownership of Model Training and Content Generated by a Generative Artificial Intelligence Tool, engrossed version (4 July 2025) <<https://arkleg.state.ar.us/Home/FTPDocument?path=%2FBills%2F2025R%2FPublic%2FHB1876.pdf>> accessed 29 July 2025.

⁵⁶ *ibid*, section 1, para 18-4-101 (a) (1).

⁵⁷ *ibid*, section 1, para 18-4-101 (a) (2).

ownership belongs to the employer. Importantly, the Act explicitly states that if the content generated by AI infringes upon pre-existing copyrights or other IP rights, the ownership over such content is not granted.

Although this legislative initiative may seem reasonable on the one hand, on the other, it is not justified from the perspective of copyright and civil rights legislation, as it has created a new object of civil rights without a corresponding set of related rights and obligations.

In September 2024, California enacted the California AI Transparency Act (Senate Bill 942)⁵⁸, which introduces requirements for generative AI providers. Starting from January 2026, ‘covered providers’, defined as entities operating a generative AI system with over one million monthly users accessible in California, must provide users with a free AI detection tool capable of identifying whether image, video, or audio content was created or altered by their system. Providers are also required to offer users a manifest disclosure (a visible label clearly stating ‘AI-generated’) and embed a latent disclosure (machine-readable origin information such as provider name, model version, timestamp, and unique identifier) directly into AI-created output.

At present, the California State Legislature is reviewing Assembly Bill 412 (AB 412)⁵⁹, a legislative proposal aimed at enhancing transparency and accountability in the process of using copyrighted content for training generative artificial intelligence models. The bill would require developers who make generative AI systems available in California to disclose any known copyrighted materials used during training, including available registration or indexing information, and to publish this documentation on their websites. It also demands that developers provide a mechanism for copyright holders to inquire whether their specific works were used in training datasets. Upon receiving such a request submitted with a digital fingerprint of the work, the developer must respond within 30 days with a good-faith assessment and, if applicable, detailed information about the use. Each day of failure to respond constitutes a separate legal violation. The bill allows rightsholders to seek civil remedies for non-compliance and exempts developers whose models were trained exclusively on public domain content, proprietary data, or materials made publicly available free of charge. Thus, if the AB 412 adopted, it would become a significant development in state-level copyright governance in the AI context.

⁵⁸ California, Senate Bill 942 California AI Transparency Act, 2023-2024 Regular Session, California Legislature <https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942> accessed 30 July 2025.

⁵⁹ California, Assembly Bill 412 (2025-2026 Regular Session), Generative Artificial Intelligence: Training Data: Copyrighted Materials AB-412 (amended 7 May 2025) <https://leginfo.ca.gov/faces/billHistoryClient.xhtml?bill_id=202520260AB412> accessed 30 July 2025.

Negative commentary and expert criticism of California's AB 412 have already appeared. For example, one of the critics argued that the bill imposes excessive and impractical compliance burdens on AI developers, particularly small and mid-sized companies that may lack the resources to document training datasets and respond to rightsholder inquiries⁶⁰. The bill was also described as premature, provided that court rulings regarding the fair use of copyrighted content in AI training are still pending⁶¹. It is also noted that the enactment of state-level legislation at this stage could create legal inconsistencies and interfere with the development of coherent federal case law.

In 2024, California introduced Assembly Bill 2013,⁶² which includes provisions related to the disclosure of training data sources used in the development of artificial intelligence systems. Under the bill, developers would be required to document and make available information about the origin, nature, and composition of the data used to train their AI models. This includes identifying whether datasets contain copyrighted materials, sensitive personal information, or other types of protected content. The bill instructs the California Office of Data and Innovation to issue detailed regulations specifying the format and scope of such disclosures.

3.3.4. Legislation perspectives

On 16 June 2025, the Congressional Research Service published a legal sidebar 'Generative Artificial Intelligence and Copyright Law'.⁶³ In this document, the CRS noted that, despite growing public debate and several legally significant lawsuits concerning the use of copyrighted content for AI training and the copyrightability of AI-generated outputs, Congress has so far chosen to refrain from making changes to existing laws.

The report highlights several reasons for this legislative approach. Most importantly, the relevant case law is still developing, with multiple federal courts currently considering foundational questions, such as whether the unauthorized use of copyrighted works to train AI systems constitutes fair use and whether AI-generated outputs can be protected by copyright in the absence of human authorship. The CRS explains that these disputes raise novel legal issues that lack settled precedent, and emphasises that judicial interpretation of existing law is expected to play a crucial role in defining the scope of permissible AI-related uses. As the report states: 'As the courts decide

⁶⁰ John Villasenor, 'The problems with California's pending AI copyright legislation' (2025) *the Brookings Institution* <<https://www.brookings.edu/articles/the-problems-with-californias-pending-ai-copyright-legislation/>> accessed 30 July 2025.

⁶¹ *ibid.*

⁶² California, Assembly Bill 2013, Generative artificial intelligence: training data transparency AB-2013, chapter 817 (2024) <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013> accessed 30 July 2025.

⁶³ Congressional Research Service, Generative Artificial Intelligence and Copyright Law, LSB10922 (2025) <<https://www.congress.gov/crs-product/LSB10922>> accessed 31 July 2025.

cases involving generative AI, they may be able to provide greater guidance and predictability in this area. Based on the outcomes of these cases, Congress may reassess whether legislation is needed.’

In this context, Congress has chosen to maintain a monitoring position, acknowledging the complexity of the issues and the need for further legal and technological clarity before pursuing statutory reform if it is necessary.

3.3.5. The role of the U.S. Copyright Office and its influence on the application of copyright law

The legal framework for AI and copyright is shaped not only by statutes and case law, but also by the institutional actors. One of the most influential among them in the US is the U.S. Copyright Office, which actively contributes to the development of copyright law through policy guidance, expert opinions, and public engagement.

Under 17 U.S. Code § 701, the U.S. Copyright Office is responsible for administering key aspects of copyright law in the United States. Its duties include managing the registration and recordation systems, advising Congress, and conducting public studies. The U.S. Copyright Office also has the authority to issue legally binding regulations that clarify and implement the provisions of the U.S. Copyright Act. These regulations are primarily set out in title 37 of the Code of Federal Regulations; they are legally binding and have the force of law within the scope of the U.S. Copyright Act.

The U.S. Copyright Office also provides information and assistance to courts and executive branch agencies⁶⁴. Its mission is to promote ‘creativity and free expression by administering the nation’s copyright laws and by providing impartial, expert advice on copyright law and policy for the benefit of all’.

The U.S. Copyright Office is responsible for registering copyright claims and maintaining the public record of registered works. In carrying out these duties, the U.S. Copyright Office regularly evaluates whether submitted works meet the requirements for copyright protection. This involves the interpretation of relevant legislation and case law, giving the U.S. Copyright Office significant expertise in the field of copyrightability.

⁶⁴ See, U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (n 9), about the U.S. Copyright Office.

On November 3, 2018, the U.S. Copyright Office received an application to register a work titled ‘A Recent Entrance to Paradise’, submitted by Stephen Thaler. The application identified the ‘Creativity Machine’ as the author, and described the work as ‘autonomously created by a computer algorithm running on a machine’. Thaler claimed the work was made for hire and sought registration in his own name as the owner of the machine.

After the U.S. Copyright Office refused to register the work on the grounds that it lacked human authorship, Thaler filed a lawsuit challenging the decision. This case became one of the well-known and widely discussed in the context of copyright and artificial intelligence, as it directly addressed whether a work created solely by a machine without any human involvement can be eligible for copyright protection.

Thus, although the U.S. Copyright Office is responsible for assessing and registering copyright claims, its decisions can be challenged in court. As noted in the CRS’s Legal Sidebar⁶⁵ ‘while the Copyright Office notes that courts sometimes give weight to the office’s experience and expertise, courts are not legally bound to adopt the office’s interpretations of the Copyright Act, such as its application of the authorship requirement to AI-assisted works’. However, as disputes over the copyrightability of AI-generated works continue to emerge, a growing amount of case law will likely develop, in which the U.S. Copyright Office and courts together will shape how copyright law adapts to the realities of artificial intelligence.

In March 2023, the U.S. Copyright Office issued official guidance titled Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence⁶⁶, clarifying its registration policy regarding works created by or with the assistance of AI.

At the same time in 2023, the U.S. Copyright Office launched a comprehensive public study on the implications of artificial intelligence for copyright law⁶⁷. This initiative began with a formal ‘Notice of Inquiry’ published in August 2023, the purpose of which was to gather opinions and suggestions from a broad range of stakeholders, including copyright holders, AI developers, legal scholars, artists, public interest groups, and industry representatives. The inquiry generated over 10,000 responses⁶⁸, reflecting wide public engagement and highlighting the complexity of the

⁶⁵ Legal Sidebar (n 63).

⁶⁶ U.S. Copyright Office, ‘Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence’, 88 FR 16190, (2023).

⁶⁷ U.S. Copyright Office, Notice of inquiry and request for comments ‘Artificial Intelligence and Copyright’, 88 FR 59942 (2023) <<https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf>> accessed 01 August 2025.

⁶⁸ U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (n 9).

issues. Based on the findings and the analysis of the input received, the U.S. Copyright Office initiated a multi-part Copyright and Artificial Intelligence report series. To date, two parts have been officially released: Part 1 (July 2024)⁶⁹, which addressed the legal treatment of digital replicas, and Part 2 (January 2025), which focused on the copyrightability of AI-generated works. A third report⁷⁰, examining the use of copyrighted materials in AI training, was published in May 2025 as a pre-release version and is expected to be finalized shortly.

Thus, as of today, the U.S. Copyright Office has established its policy on the copyrightability of AI-generated content and the use of protected works for AI training, which will be examined in detail in Chapter IV.

3.3.6. Conclusion

Thus, as of mid-2025, the United States still lacks a unified federal legal framework specifically addressing the use of copyrighted content in AI development. In this legal vacuum, many states have started to introduce their own regulations, often with very different approaches and priorities. While this reflects growing concern about the role of AI, it also creates a highly fragmented legal environment. Developers and deployers of AI systems may soon face the complex task of complying with multiple, and possibly conflicting, state laws—each with its own disclosure, consent, or ownership requirements. Given how fast both AI technologies and legislative initiatives are evolving, this could result in significant legal uncertainty and financial risks, especially for companies operating across states. Without clear federal coordination, the cost of constantly adapting to new legal obligations in dozens of jurisdictions may become a serious barrier to innovation.

In this context, the role of the U.S. Copyright Office has become increasingly central. Through updated guidance, public consultations, and its ongoing multi-part report, the U.S. Copyright Office has taken the lead in clarifying how copyright law should apply to AI-generated works and the use of protected content in AI training. However, without unified federal legislation, the United States continues to rely on administrative interpretation and fragmented state initiatives.

3.4. Comparative analysis: EU vs US legislative approaches to AI and copyright

The European Union and the United States apply different legislative approaches in their treatment of AI and copyright, as evidenced in this Chapter. The EU's approach is characterised by the early integration of AI regulations into its existing framework by the adoption of the AI Act. By contrast,

⁶⁹ U.S. Copyright Office, 'Copyright and Artificial Intelligence: Part 1: Digital Replicas', (2024), <[Copyright and Artificial Intelligence, Part 1 Digital Replicas Report](#)> accessed 11 July 2025.

⁷⁰ U.S. Copyright Office, 'Copyright and Artificial Intelligence, Part 3: Generative AI Training' (n 1).

the United States has no comprehensive federal legislative instrument specifically regulating the intersection of AI and copyright, relying instead on administrative interpretation, evolving case law, and state-level initiatives.

In the EU, the primary legislative framework relevant to AI and copyright emerged not from creating a wholly new copyright regime, but from adapting existing law, most notably the DSM Directive's TDM exceptions. The AI Act includes copyright compliance obligations, imposing enforceable transparency requirements, opt-out respect, and documentation duties on providers of general-purpose AI models. The EU's strategy reflects a harmonised model, ensuring that the same obligations apply uniformly across Member States. As discussed further in Chapter IV, the application of TDM exceptions to AI training may not have been the most successful regulatory solution. Nevertheless, the European Union is making efforts to at least partially regulate this area, seeking to find a balance between protecting rightsholders and supporting technological development.

In the United States, there is still no federal law that specifically regulates the intersection of copyright and AI. Instead, sector-specific statutes and executive orders have addressed AI more generally, while copyright questions are left to judicial interpretation and administrative guidance from the U.S. Copyright Office. The current result is a decentralized, fragmented framework, with substantive rules emerging through litigation and agency policy rather than a unified legislative initiative.

EU legislative output is driven primarily by the European Commission and the European Parliament, with the European AI Office established to ensure uniform implementation, develop compliance guidance, and support harmonisation across Member States.

In the US, at the moment, the U.S. Copyright Office plays a more substantive role in shaping AI copyright policy than any federal legislative body. By establishing registration rules for AI-generated outputs, public consultations, and detailed policy reports, the U.S. Copyright Office has set interpretative baselines for copyrightability and the use of protected works in AI training. However, its interpretations are not binding on courts, creating potential divergence between administrative guidance and judicial outcomes.

The EU's AI Act imposes a clear transparency obligation: AI providers must publish detailed summaries of training data sources, enabling rights holders to verify potential infringements. These obligations are precise, structured, and legally enforceable across all Member States.

In the US, transparency requirements are not federally mandated but are emerging at the state level. California's pending AB 412, for example, would impose dataset disclosure and rightsholder inquiry mechanisms similar to EU transparency provisions, but only within state jurisdiction. Such initiatives risk producing conflicting obligations varying by state, with possible compliance problems for multi-state or national AI providers.

While the EU framework focuses primarily on rights compliance and does not establish a new authorship model for AI-generated works, US state-level laws have begun to define the rules of ownership directly. Thus, Arkansas Act 927 assigns ownership to the individual providing input or data. This contradicts the current positions of legislators and courts in both the EU and the US, which reject protection for works lacking human authorship. This creates a potential conflict between state-level proprietary claims and federal copyright doctrine.

The EU's centralised model reduces internal legal fragmentation, although it risks imposing burdensome compliance costs and potentially slowing market entry for some developers. The U.S. decentralised model, on the one hand, fosters flexibility and avoids premature statutory amendments without deeper investigation and judicial decisions; on the other hand, it generates significant uncertainty for stakeholders, who must navigate both inconsistent state laws and unsettled federal jurisprudence. Without federal harmonisation, state-by-state divergence could raise compliance costs and complicate AI deployment.

Thereby, the EU and the US approaches reveal a structural contrast: the EU favors a harmonised, pre-emptive legislative framework with embedded copyright safeguards, while the US currently relies on administrative interpretation and awaits court decisions, which is justified for the common law system.

IV. Challenges and Complex Issues

4.1. Copyrightability of the AI-generated content

4.1.1. Introduction

One of the most interesting and controversial issues at the intersection of artificial intelligence and copyright law concerns the legal status of AI-generated content. Fundamental questions are raised as AI systems create more and more creative output that simulates human-authored works, such as text, images, music, and other media: can such output be protected by copyright? If so, who should be recognised as the rightful holder of the copyrights? These questions are not just theoretical—they have a significant impact on ownership, licensing, and enforcement in creative industries, as well as on the broader functioning of copyright systems designed to incentivise and reward human creativity. Both the EU and the US are actively debating these issues, and as discussed above, there are currently no legislative changes that specifically address this problem. Therefore, any legal assessment must rely on the current EU and US copyright framework. However, since the nature of AI-generated outputs is not completely clear and AI technologies are still in the early stages of development, there is ongoing uncertainty about whether the existing legal rules provide sufficiently clear answers. As a result, lawmakers, scholars, and industry stakeholders continue to discuss whether changes to the legal framework may eventually be necessary or whether the current framework is sufficient for answering these questions.

4.1.2. Copyrightability under the EU legal framework

In order to determine whether AI-generated output qualifies for copyright protection, it is first necessary to identify the criteria that define a protectable work under EU copyright law.

Several comprehensive studies have been conducted within the European Union regarding the issue of copyright protection for AI-generated works⁷¹. In these studies, the authors consistently conclude that only a human being can be recognised as the author of a work. This position is based on several key arguments.

The EU directives that harmonise copyright law across the Member States do not provide an exhaustive list of protected subject matter, nor do they offer a definition of the terms ‘works’⁷² or ‘author’ with limited exceptions⁷³. However, Article 9(1) of the TRIPS states that ‘Members shall

⁷¹ See e.g., European Commission, Trends and Developments in Artificial Intelligence (n 31), Lucchi (n 14), Oleksandr Bulayenko and others ‘AI Music Outputs: Challenges to the Copyright Legal Framework’, reCreating Europe Report (2022) <<https://zenodo.org/records/6405796#.YkbhSihBw2w>> accessed 15 July 2025.

⁷² AI Music Outputs (n 71), 28.

⁷³ See footnote 12 for a detailed explanation.

comply with Articles 1 through 21 of the Berne Convention (1971) and the Appendix thereto⁷⁴. Thus, the Berne Convention⁷⁵ is incorporated into the EU legal framework, and the EU directives regulating copyright matters make explicit reference to it⁷⁶.

The Berne Convention repeatedly refers to the ‘author’ as the creator of a work and the beneficiary of moral and economic rights⁷⁷. Article 2 of the Berne Convention provides that it protects ‘literary and artistic works’ and sets out a non-exhaustive list of categories of such works. Over time, CJEU has played a central role in clarifying the concept of a ‘work’ and the ‘author’ under EU law. Landmark decisions on the matter include *Infopaq International A/S v Danske Dagblades Forening (Infopaq)*, where the Court held that ‘copyright...is liable to apply only in relation to a subject-matter which is original in the sense that it is its author’s own intellectual creation’⁷⁸; and *Painer v. Standard Verlags GmbH (Painer)*, where the Court confirmed that the work is protected by copyright if it is the result of ‘intellectual creation of the author reflecting his personality and expressing his free and creative choices’⁷⁹; and other decisions, where the Court supported the same position⁸⁰.

From the decisions mentioned above, it is clear that authorship under EU copyright law is directly linked to creativity, personality, and the exercise of free and creative choices—attributes that can be associated exclusively with a human being.

As further clarified by the case law of the CJEU, a fundamental requirement for a work to be eligible for copyright protection is originality, meaning that the work must be created by the author and reflect the author’s own intellectual creation⁸¹. ‘The requirement of originality does not imply artistic merit or aesthetic quality’⁸², but it does exclude purely functional, technical, or informational content—especially if it leaves little or no room for creative expression⁸³. For instance, purely factual military reports (as in *Funke Medien NRW GmbH v Bundesrepublik Deutschland*⁸⁴ (*Funke Medien*)) are not protected as ‘works’ because they lack human authorship

⁷⁴ Agreement on Trade-Related Aspects of Intellectual Property Rights (15 April 1994) 1869, UNTS 299, art. 9(1).

⁷⁵ Berne Convention (n 22).

⁷⁶ See, e.g., InfoSoc Directive (n 34), recital 15; Term Directive (n 34), art. 1.

⁷⁷ European Commission, Trends and Developments in Artificial Intelligence (n 31), 68.

⁷⁸ C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para 37.

⁷⁹ C-145/10 *Painer v Standard Verlags GmbH and Others* [2011] ECLI:EU:C:2013:138, para 99.

⁸⁰ See, e.g., C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* [2019] ECLI:EU:C:2019:623; C-683/17 *Cofemel – Sociedade de Vestuário SA v G-Star Raw CV* [2019] ECLI:EU:C:2019:721.

⁸¹ C-403/08 and C-429/08 *Football Association Premier League Ltd and Others v QC Leisure and Others* [2011] EU:C:2011:631, para 97, Case C-310/17 *Levola Hengelo BV v Smilde Foods BV* [2018] ECLI:EU:C:2018:899, para 37.

⁸² See, European Commission, Trends and Developments in Artificial Intelligence (n 31), 71.

⁸³ *ibid*, 78.

⁸⁴ *Funke Medien NRW GmbH v Bundesrepublik Deutschland* (n 80).

and creative intent⁸⁵. The CJEU, however, has not set a precise minimum threshold of creativity. Early rulings, such as *Infopaq*, suggest that even minimal creative input, like the selection of eleven words, may be sufficient. Similarly, in *Painer*, the Court implied that even standardised portrait photographs can meet the originality requirement if they result from personal and creative decisions⁸⁶. Thus, even a minimal level of creativity may be sufficient to justify copyright protection.

In addition to the requirement of originality, several other important findings derived from the case law of the CJEU help to assess whether AI-generated content may qualify for copyright protection⁸⁷. These include:

1. Creative choice over economic or technical investment

EU copyright law focuses on the creative process rather than the effort or investment involved. In *Football Dataco*⁸⁸, the Court rejected the idea that ‘skill and labour’ alone are sufficient for originality, ruling that copyright cannot be granted merely on the basis of economic or technical input.

2. Functional constraints limit originality; ideas are not protectable

In the *Brompton Bicycle* case⁸⁹, the CJEU held that copyright protection only extends to works whose creation is not dictated solely by technical considerations. The Court emphasized that where the expression of a work is entirely determined by its technical function, the criterion of originality is not fulfilled. The Court reaffirmed that, under Article 2 of the WIPO Copyright Treaty⁹⁰, copyright protection does not extend to ideas, and that protecting ideas would lead to their monopolisation, to the detriment of technical progress and industrial development.

3. Ideas are not protected, but their creative combination might be

While ideas in themselves are not protected, the *Painer* decision illustrates that a combination of otherwise unprotectable elements — such as pose, lighting, angle, or background in photography — may, when shaped by creative choice, lead to an original work. This is crucial for assessing AI-assisted outputs, where the role of the human often lies in conceptual design and choices rather than in creating the expressive form.

⁸⁵ *Levola Hengelo BV v Smilde Foods BV* (n 81).

⁸⁶ See, European Commission, Trends and Developments in Artificial Intelligence (n 31), 74.

⁸⁷ *ibid*, 71-75.

⁸⁸ C-604/10 *Football Dataco Ltd v Yahoo! UK Ltd* [2012] ECLI:EU:C:2012:115.

⁸⁹ C-833/18 *SI and Brompton Bicycle Ltd v Chedech/Get2Get* [2020] ECLI:EU:C:2020:461, paras 27, 33.

⁹⁰ WIPO Copyright Treaty (adopted 20 December 1996) 2186 UNTS 121, art. 2.

4. Creativity may occur at any stage

The CJEU recognises that creative choices may be made not only during the execution but also in the preparatory or conceptual stages of a work⁹¹. This insight opens the door for recognising originality in AI-assisted outputs, provided the human author made key creative decisions during conception or planning.

5. Expression of creative choice is essential

A protected work must not only be the result of creative activity, but that creativity must also be expressed in the work itself. In *Infopaq* and *Painer*, the Court made clear that expression — in a form that is identifiable with sufficient precision and objectivity — is a necessary condition for copyright protection. In *Levola Hengelo*, the Court ruled that the output must be identifiable with sufficient clarity and objectivity. This excludes purely sensory or subjective experiences (such as taste), reinforcing that copyright only applies where there is a clearly defined expression.

6. General authorial intent could be considered

Although authorial intent and creative conception are generally regarded as part of the underlying idea, which is not protected under copyright law, there is an opinion suggesting that they may nonetheless be taken into account when assessing originality and the author's creative choices⁹². In his Opinion in the joined cases *Mio* and *Konektra*⁹³, Advocate General Szpunar clarified that while authorial intent is not decisive in determining originality, it may be taken into account by courts when assessing whether a work qualifies for copyright protection. The key requirement remains that the work must objectively express the author's free and creative choices, reflecting their personality. However, if credible evidence of the creator's intentions is available, courts may consider it as supporting evidence, provided that those intentions are also perceptible in the work itself.

Thus, based on the current EU legal framework and the case law of the CJEU, the copyrightability of AI-generated outputs must be assessed by reference to established criteria that define a protectable work. The core requirement is a human authorship and originality. This implies that the work must reflect free and creative choices made by a human author, and that these choices must be objectively expressed in the resulting output.

⁹¹ *Painer* (n 79), para 91.

⁹² See, European Commission, Trends and Developments in Artificial Intelligence (n 31), 75.

⁹³ Opinion of AG Szpunar, joined Cases C-580/23 and C-795/23 *Mio AB and Others v Galleri Mikael & Thomas Asplund Aktiebolag; konektra GmbH, LN v USM U. Schärer Söhne AG* [2025] ECLI:EU:C:2025:330.

However, the jurisprudence of the CJEU offers further guidance that may be relevant in evaluating the copyrightability of AI-assisted outputs, particularly where a human has played a role in shaping the result. For instance, it has been established that only creativity should be taken into account, not mere investment or labor; that creative choices may occur at various stages of the process, including the preparatory or conceptual phase; and that some commentators suggest authorial intent, while not decisive, might nonetheless be considered when it is reflected in the objectively perceptible features of the work. These interpretative principles may prove useful in future analysis of copyright questions related to AI-generated content.

4.1.3. Copyrightability of AI-generated output under the EU legal framework

There are generally two categories of AI-generated output. The first includes output generated autonomously by an AI system, without any human exercising creative judgment at any stage of the process. The second refers to output produced with the assistance of AI, but resulting from human creative input—such as formulating the prompt, shaping the general concept of the desired output, refining or modifying the prompt to influence the result, or revising and editing the AI-generated content.

As for the first type of output, as discussed above, it does not meet the definition of a ‘work’ under EU copyright law, lacks an identifiable author, and therefore falls outside the scope of copyright protection. This view is widely supported in legal scholarship and reflected in official reports and statements issued by public authorities⁹⁴. In the absence of any legal provisions granting ownership rights over such output, it is generally understood that these materials fall into the public domain. This has significant legal and commercial consequences: individuals and companies that invest in AI technologies or use AI systems to generate creative content cannot claim exclusive or proprietary rights in the absence of human authorship⁹⁵. However, for some AI-assisted outputs, alternative protection regimes may apply, such as film producer rights for audio-visual content, phonographic rights for audio data, and broadcasting rights for AI-assisted broadcasts⁹⁶. These regimes offer an alternative form of protection for the creators of such outputs.

By contrast, the second category of AI-assisted output may be eligible for copyright protection—provided that a human contributes creatively to the process. However, this determination must be made on a case-by-case basis, as the degree of human involvement may vary substantially. In some cases, AI is merely used as a tool in the execution of a human-created work, much like Adobe

⁹⁴ See e.g., European Commission, Trends and Developments in Artificial Intelligence (n 31), 83; Resolution on artificial intelligence in education, culture and the audiovisual sector (n 32).

⁹⁵ See, Lucchi (n 14), 93.

⁹⁶ See, European Commission, Trends and Developments in Artificial Intelligence (n 31), 117.

Photoshop is used by photographers for editing images, or digital audio workstations such as Ableton Live or Logic Pro are used by musicians to compose and produce music. In others, human input may be minimal, limited to providing a basic prompt, while the AI system generates the output autonomously.

As it is concluded in the report of the European Commission, Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework (2020)⁹⁷ (hereinafter EU AI and IP Report)

‘the current EU copyright law, as interpreted by the CJEU, leaves room for the protection of AI-assisted outputs in a wide range of creative fields. As long as the output reflects creative choices by a human being at any stage of the production process, an AI-assisted output is likely to qualify for copyright protection. This is true even if the AI system has played a significant or even predominant role in the entire creative process’⁹⁸.

The creative choices while interacting with AI ‘may occur at various levels and in various phases of the creative process: conception/preparation, execution, and finalisation’⁹⁹. While AI may perform the execution phase, sufficient human creative involvement in the other stages may result in the creation of a copyright-protected work.

In addition, the following findings of the EU AI and IP Report appear to be particularly important in the context of this analysis:

- 1) due to the ‘black box’ nature of some AI systems, individuals responsible for the conception phase may not be able to accurately predict or explain the final output of the execution phase. However, this uncertainty does not exclude that the resulting output may be qualified as a copyright-protected work, provided it remains within the bounds of the author’s general creative intent;
- 2) copyright should belong to the user who created the content, rather than to the developer of the AI system, except in cases where the author and the developer collaborate on a specific creative project, in which case co-authorship may arise;
- 3) it is not possible to make general determinations regarding the copyright status of AI-assisted outputs. In cases where creative choices cannot be clearly identified, copyright protection is unlikely to apply.

Despite the comprehensive nature of the European Commission’s study, its conclusions have not been universally accepted. For example, in his study ‘Artificial Intelligence and the Future of

⁹⁷ European Commission, Trends and Developments in Artificial Intelligence (n 31).

⁹⁸ *ibid*, 76.

⁹⁹ *ibid*.

Copyright: New Challenges for the EU', Nicola Lucchi raises important questions¹⁰⁰. He highlights that EU copyright law is based on the concept of human authorship, regardless of the complexity or human-like qualities of the outputs autonomously generated by AI. This creates a grey area in which valuable content may fall outside the scope of legal protection, raising issues related to ownership, liability, and reuse.

The European Parliament addressed this issue in its Resolution on artificial intelligence in education, culture and the audiovisual sector¹⁰¹ adopted following the publication of the European Commission's report. In the Resolution the European Parliament generally supported the approach taken in the EU AI and IP Report and reaffirmed the principle that copyright protection in the EU applies only to human-created works. In particular, it emphasised the role of the author's personality in the expression of free and creative choices, which constitute the originality of works and confirmed that 'where AI is used only as a tool to assist an author in the process of creation, the current IP framework remains applicable'¹⁰². At the same time, the Parliament acknowledged the growing relevance of AI-generated content, suggested that AI-generated content 'must be protected under the IPR legal framework in order to encourage investment in this form of creation and improve legal certainty for citizens, businesses and...inventors'¹⁰³, and emphasized that the 'ownership of rights, if any, should only be assigned to natural or legal persons that created the work'¹⁰⁴.

Thus, the text of the Resolution indicates that the European Parliament draws a clear distinction between authorship and ownership in relation to AI-generated content. While the principle that only a human can be considered an author is generally undisputed, the question of ownership remains open and legally uncertain. The Parliament considers that fully autonomous AI-generated works may not meet the originality threshold required for copyright protection, as originality is linked to the personality of a natural person. However, it suggests that such results may still need to be protected within the broader intellectual property rights system in order to provide legal certainty, support innovation, and encourage continued investment in AI technologies.

4.1.4. Copyrightability of AI-generated outputs under the US legal framework

The question of how copyright law should respond to the development of artificial intelligence has been the subject of widespread and ongoing debate in the United States. Although there have not

¹⁰⁰ Lucchi, (n 14), 27.

¹⁰¹ Resolution on artificial intelligence in education, culture and the audiovisual sector, (n 32).

¹⁰² *ibid*, paras 13, 14.

¹⁰³ *ibid*, para 15.

¹⁰⁴ *ibid*.

yet been major changes to the law, both Congress and the U.S. Copyright Office are actively working to clarify their current positions. The U.S. Copyright Office has published detailed guidelines and reports, most recently about digital replicas and whether AI-generated works can be protected by copyright. Congress has released Legal Sidebar¹⁰⁵ to express its position and concerns on the matter. For now, policymakers are taking a careful, wait-and-see approach, monitoring how case law develops and how public and professional opinion continues to evolve. Stakeholders are actively participating in this discussion—lawyers, academics, industry professionals, artists, and technology experts. The debate is also supported by research studies, university conferences, seminars, and articles in major media. This multi-level interaction reflects the growing urgency of determining how copyright law should adapt to technological advancements while maintaining its fundamental purpose of encouraging human creativity.

In the United States, as in the European Union, the conditions for copyrightability of works are defined on the basis of statutory law and a substantial body of case law that has clarified the interpretation of legal provisions. The major role played by the U.S. Copyright Office.

As discussed in Chapter III, in March 2023, the U.S. Copyright Office issued ‘Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence’¹⁰⁶, clarifying its registration policy regarding works created by or with the assistance of AI. According to the U.S. Copyright Office, only works that are the product of human authorship are eligible for copyright protection. Therefore, works created entirely by AI without human creative input cannot be registered. However, the U.S. Copyright Office acknowledges that copyright protection may apply to works that combine human and artificially created content, but only the human-created elements will be protected. Applicants are required to explicitly exclude the AI-generated portions of outputs in their registration applications. The U.S. Copyright Office emphasised that merely providing prompts to an AI system is insufficient to claim authorship, as the expressive decisions are made by the AI, not the user.

In the report ‘Copyright and Artificial Intelligence, Part 2: Copyrightability’ (hereinafter ‘the Copyrightability Report’), the U.S. Copyright Office examined in detail whether existing legal doctrines are sufficient to resolve questions of copyrightability in the context of AI-generated content. In this report, the U.S. Copyright Office emphasises that the principal requirement for copyrightability—under both statutory law and established case law—is human authorship¹⁰⁷. It

¹⁰⁵ Legal Sidebar (n 63).

¹⁰⁶ Copyright Registration Guidance (n 66).

¹⁰⁷ U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (n 9), 7.

refers to the number of cases where courts denied copyright to nonhuman creators such as animals¹⁰⁸ and supernatural entities¹⁰⁹.

Moreover, as it was noted in Chapter III, the U.S. Copyright Office itself already participates in the court hearings on the matter after it denied registering a piece of visual art, which was claimed as created by an AI. The applicant, Stephen Thaler, filed a claim against the U.S. Copyright Office to appeal this decision. In 2023, a federal district court ruled against him, affirming that human authorship is essential to a valid copyright claim and that the machine cannot be admitted as an author of the generated output¹¹⁰. As confirmed by the D.C. Circuit Court of Appeals in March 2025¹¹¹, although the Copyright Act¹¹² does not explicitly define the term ‘author’ ‘many of the Copyright Act’s provisions make sense only if an author is a human being’, therefore, ‘the best reading of the Copyright Act is that human authorship is required for registration’¹¹³. The Court emphasised that the Copyright Act implicitly refers to human beings, as evidenced by its context—particularly provisions concerning the duration of copyright protection based on the author’s lifetime, inheritance rules, the requirement of a signature to transfer copyright ownership, and the extension of protection to unpublished works regardless of the author’s nationality or domicile.

Based on these arguments, the U.S. Copyright Office concluded in its Copyrightability Report that only a human can be considered the author of a work.

The next issue examined in the Copyrightability Report was the copyrightability of works created with the assistance of AI. On this matter, the U.S. Copyright Office reaffirmed the position it had already stated in its earlier Guidance¹¹⁴: if a work contains more than a minimum amount of AI-generated material, only the portions that reflect human creative input are eligible for copyright protection. ‘Where AI merely assists an author in the creative process, its use does not change the copyrightability of the output. At the other extreme, if content is entirely generated by AI, it cannot be protected by copyright. Between these boundaries, various forms and combinations of human contributions can be involved in producing AI outputs’¹¹⁵. The U.S. Copyright Office clarified that

¹⁰⁸ This refers to the case *Naruto v. Slater* 888 F.3d 418 (9th Cir. 2018), in which People for the Ethical Treatment of Animals filed a lawsuit requesting that copyright ownership be assigned to the macaque. The court ultimately held that animals do not have statutory standing to sue under the Copyright Act.

¹⁰⁹ In *Urantia Foundation v. Maaherra* 114 F.3d 955 (9th Cir. 1997) the Ninth Circuit held that a book allegedly transmitted by celestial beings could qualify for copyright protection, provided that there was sufficient human selection, arrangement, or other creative contribution in its final form. The court reaffirmed that copyright does not protect divine or supernatural authorship but may extend to the human expression of such content if it reflects the author’s own intellectual creation.

¹¹⁰ *Thaler v. Perlmutter*, No 1:22-cv-01564-BAH (D.D.C 2023).

¹¹¹ *Thaler v. Perlmutter*, No. 23-5233 (D.C. Cir. Mar. 18, 2025) <[23-5233.pdf](#)> accessed 2 August 2025.

¹¹² Copyright Act of 1976, 17 USC §§ 101–810.

¹¹³ *Thaler* (DC Cir) (n 111), 10.

¹¹⁴ Copyright Registration Guidance (n 66).

¹¹⁵ U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (n 9), 2.

when a human contributes original, copyrightable content that is perceptible in the final AI-generated output, that individual will be considered the author of at least that portion of the work. In such cases, copyright protection would apply to the human-authored elements, similarly to how derivative works are protected, limited to the material added by the later author. Additionally, copyright may extend to the selection, coordination, and arrangement of both human and AI-generated components, although it would not cover the AI-generated elements in isolation¹¹⁶.

As regards prompts, the U.S. Copyright Office said that they may express a user's ideas, but they do not dictate the specific expressive elements of the output. 'The fact that identical prompts can generate multiple different outputs further indicates a lack of human control'¹¹⁷. Thus the U.S. Copyright Office concluded that, based on currently available technology, prompts alone do not give users sufficient creative control to qualify as authors of AI-generated outputs.

The U.S. Copyright Office has also addressed the question of whether legislative changes are needed. In its opinion, such changes are currently unnecessary. Since each work involving AI must be assessed individually, it would be difficult for legislation to offer more clarity than what case-by-case judicial analysis can provide. Therefore, as concluded by the U.S. Copyright Office, the most appropriate course of action at this stage is to observe how the case law develops¹¹⁸.

Thus, by issuing its official Registration Guidance in March 2023 and through its ongoing Copyright and Artificial Intelligence initiative, the U.S. Copyright Office has formally confirmed that it currently accepts and reviews applications for registration of content created with the assistance of AI. Moreover, the U.S. Copyright Office has already registered hundreds of works¹¹⁹ in which the applicant demonstrated creative choices, selection, coordination, or modification of AI-generated material, as long as those elements reflected a sufficient level of human creativity and originality.

More recently, the US Congress also addressed the issues under consideration. In July 2025, it published a Legal Sidebar¹²⁰, which provides the current legal understanding of whether copyright protection may be afforded to AI-generated outputs. The document has referred to the same court decisions and the position of the U.S. Copyright Office and generally aligns with the U.S. Copyright Office's conclusions.

¹¹⁶ U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (n 9), 24.

¹¹⁷ *ibid*, 20.

¹¹⁸ *ibid*, 40.

¹¹⁹ *ibid*, 3.

¹²⁰ Legal Sidebar, (n 63).

At the same time, legal scholars are proposing other theories regarding the ownership and copyrightability of AI-generated outputs. For example, Mark Lemley, a prominent US legal scholar and professor at Stanford Law School specializing in intellectual property, proposes a provocative idea that the challenge of copyrightability of AI-generated outputs may lead to a fundamental transformation of copyright law¹²¹. He argues that traditionally, copyright protects the specific form of expression of a creative idea made by a human author; however, in the case of generative AI, the form is produced autonomously by the machine, while human input may be limited to a prompt. Since both legislation and case law recognise only human authorship, such outputs currently fall outside the scope of protection. Lemley observes that, increasingly, human creativity lies in asking the right questions rather than providing the answers. This shift suggests a possible future in which copyright protection focuses not on the AI-generated output itself, but on the structure and originality of the prompts, assuming they contain sufficient creative input. Lemley mentions that although the U.S. Copyright Office has acknowledged this approach in theory, in practice, it has rejected applications even where significant prompt engineering was involved. He argues that this ‘prompt-based’ model reverses the current logic of copyright, potentially requiring a rethinking of both the concept of authorship and the traditional infringement tests, which rely on similarities in expression. As generative AI assumes the role of producing the expressive content, human contribution becomes increasingly limited to unprotectable ideas and concepts, thus turning the foundations of copyright law upside down.

4.1.5. Observations and conclusions on copyrightability of AI-generated output

The legislation and case law in both the European Union and the United States generally agree that only a human can be considered the author of a work. As demonstrated in academic studies, literature, and court decisions, this follows clearly from the logic of existing copyright rules, even if the laws do not explicitly state that authors must be human. On this point, there is little doubt.

When it comes to outputs created with some level of human creative involvement—such as detailed, thoughtful prompts or significant editing after generation—the situation becomes more complex. If the AI-generated material has been creatively modified after its creation, it should be treated similarly to derivative works. In such cases, the rules are relatively clear.

However, when the output is used in its original AI-generated form, or when changes made to it do not involve sufficient human creativity to qualify as a new work, the legal status is much less certain. In both the EU and the US, there is a presumption that a work is protected by copyright from the moment it is created, without the need for registration. This means that the protection

¹²¹ See, Lemley, ‘Generative AI’ (n 13).

applies automatically to the expression of a creative idea, and when we look at a painting, a video, or a book, we can assume it is protected. In this regard, Mark Lemley is right – the form and expression will no longer be evidence of a protected work. If copyright protection depends on whether a prompt was creative enough, we will never know whether a particular work is protected or it is in the public domain. But who would take the responsibility to decide whether the prompt is creative enough? Every person who writes a prompt will likely believe it is creative, while those who want to use the resulting content may argue the opposite. In this case, the courts would have to decide. That would significantly increase the number of legal disputes and take years before any reliable criteria for evaluating prompts and notifying the public could be developed, if such criteria can even be developed at all. In the US, the U.S. Copyright Office would likely play the initial role of deciding which works qualify for protection, but it is unclear how much content would be submitted for registration and whether the U.S. Copyright Office could realistically handle the volume. And many rejected applications will likely lead to litigation — a trend that has already begun. For all these reasons, granting authorship based on the creativity of a prompt does not seem to be a workable approach.

At the same time, the clarification of the nature of the authorship does not resolve the concerns raised by scholars such as Lucchi and by the European Parliament in its Resolution on artificial intelligence in education, culture and the audiovisual sector¹²² on the ownership of AI-produced content, which is not eligible for copyright protection. The fact that AI is used as a tool for producing the content and it forms the expression does not mean that the resulting output belongs to no one. Such content may involve significant investment of time, effort, and resources—even if not necessarily of a creative nature. Thus, if AI-generated content is not protected by copyright, or if its protection is uncertain, then most probably it represents a new type of civil rights object. It may require its own legal rules and protection regime. The existing concept of authorship no longer fits the realities of how content is produced with generative AI, and it cannot effectively support both innovation and the protection of rights. A new legal category may therefore be needed.

There is no need to fundamentally change copyright law, as it continues to serve the needs of authors and effectively protect their interests. Even in the time of widespread AI-generated content, human-created works resulting from genuine creative effort will remain, and they should continue to enjoy copyright protection, which has proven to be a reliable and effective legal tool. As for AI-

¹²² Resolution on artificial intelligence in education, culture and the audiovisual sector, (n 32).

generated outputs produced without meaningful human creative involvement, it may be more appropriate to establish a separate legal regime reflecting their unique characteristics.

4.2. The use of copyrighted works for AI training

This part of the Chapter examines how the use of copyright-protected works for training artificial intelligence systems is addressed in the European Union and the United States. While the EU has chosen to regulate this area primarily through the TDM exceptions in the DSM Directive, further reinforced by provisions of the AI Act, the practical application of these rules to generative AI training has raised significant legal and policy questions. In the United States, there is no equivalent statutory framework; instead, the permissibility of AI training is evaluated under the fair use doctrine, resulting in a case-by-case approach shaped by court decisions and guidance from the U.S. Copyright Office. The following sections analyse the main legal challenges, interpretative debates, and early judicial decisions in both jurisdictions, highlighting unresolved issues and emerging trends in the regulation of AI training.

4.2.1. Can copyrighted works be used to train AI? Legal challenges under the EU copyright framework

As discussed in Chapter III, the AI Act extends the application of Articles 3 and 4 of the DSM Directive to the context of AI training. At first glance, the legal framework provides clarity: Article 4 of the DSM Directive, in conjunction with Article 53(1)(c) of the AI Act, establishes rules for the use of copyright-protected materials for text and data mining, including to train artificial intelligence, provided that rightsholders have not opted out. Thus, if a rightsholder has expressly reserved their rights, AI developers must obtain authorisation before using the content for training. However, this seemingly clear provision has not resolved all legal uncertainties. Ongoing academic and policy debates raise concerns as to whether Article 4 is fully suitable for AI training purposes and which challenges continue to arise or remain unresolved in its practical application.

4.2.1.1. Machine-readable format

One of the practical issues in applying Article 4 of the DSM Directive is that the legislator did not define what constitutes a machine-readable format, resulting in legal uncertainty as to how rightsholders' reservations of rights should be implemented in practice, particularly in the absence of commonly accepted technical standards.

Machine-readable means refer to technical mechanisms that allow rightsholders to communicate reservations of rights in a way that can be automatically recognised by web crawlers and data-

mining bots. These include the use of the robots.txt file placed at the root of a website to disallow access to specific user agents¹²³, meta tags in HTML, and HTTP headers¹²⁴. For example, in 2023, OpenAI officially announced that its web crawler, GPTBot, would respect instructions set out in websites' robots.txt files¹²⁵.

In addition to established technical tools such as the robots.txt file and meta tags, a number of new, informal practices have recently emerged that aim to limit the use of website content for AI training. The most notable examples are the machine-readable directives `noai` and `noimageai`¹²⁶, which website owners can include in their robots.txt files to indicate to automated web crawlers that their content should not be used to train artificial intelligence models. Although these directives are not part of any official technical standards or legal frameworks, they were created through community-led initiatives, especially by artists, creators, and open-source developers who were concerned about the unauthorized use of their work in large-scale AI datasets. Since then, major AI providers like OpenAI, Google, and Anthropic have begun to recognise and respect these signals as part of their voluntary self-regulation. While these measures currently lack formal legal recognition, they represent an important step toward developing practical safeguards that can be automatically read and enforced by AI crawlers, and thus complement the concept of machine-readable rights reservations referred to in Recital 18 of the DSM Directive.

While website terms and conditions can also indicate a rightsholder's reservation of rights, such notices are typically included in natural-language legal documents and are not automatically processed by web crawlers or other automated systems. This makes technical, machine-readable signals — such as entries in the robots.txt file — a more effective and reliable method for enforcing opt-outs in practice. However, the precise interpretation of what qualifies as a 'machine-readable means' under Article 4(3) of the DSM Directive remains legally unsettled. In *Robert Kneschke v. LAION e.V.*¹²⁷, the Hamburg Regional Court did not issue a final ruling on the applicability of the general TDM exception under Section 44b of the German Copyright Act, but it offered an interpretative commentary on the meaning of machine-readable disclaimers.

¹²³ DataDome, 'How to Use Your Robots.txt to (Even Partially) Block Bots From Crawling Your Site' (2025) <<https://datadome.co/bot-management-protection/blocking-with-robots-txt/>> accessed 27 July 2025.

¹²⁴ Stephan Brüggemann, 'Meta Tag Robot: Understanding Its Role in SEO Optimization' (2025) <<https://www.seocopilot.com/meta-tags/metatag-robot-understanding-its-role-in-seo-optimization>> accessed 27 July 2025.

¹²⁵ OpenAI, 'Bots on OpenAI Platform' <<https://platform.openai.com/docs/bots>> accessed 27 July 2025.

¹²⁶ DeviantArt, 'UPDATE: All Deviations Are Opted Out of AI Datasets' (2022) <<https://www.deviantart.com/team/journal/UPDATE-All-Deviations-Are-Opted-Out-of-AI-Datasets-934500371>> accessed 27 July 2025.

¹²⁷ No. 310 O 227/23, *Robert Kneschke v. LAION e.V.*, Regional Court of Hamburg, Judgment of 27 September 2024.

The court noted that the rightsholder had included a reservation of rights in natural language on the photo agency's website and acknowledged that, depending on the state of technology, such a declaration might be considered machine-readable. However, the court left open whether such a reservation would have satisfied the requirement at the time the content was reproduced in 2021. This court decision has already generated significant controversy, as it requires developers to account not only for technically coded and machine-executable formats (e.g., for crawler software), but also for opt-outs expressed in natural language. Such an interpretation expands the meaning of machine-readable and may create additional burdens for developers, particularly those not working with models capable of processing natural language¹²⁸.

4.2.1.2. AI Training is not equal to TDM

Some scholars raise a question whether it was justified to apply Article 4 of the DSM Directive to AI training, as it was included in the law for completely different purposes. They argue that this provision was intended to facilitate analytical uses of content—such as the extraction of patterns and information—rather than the synthesis and reproduction of expressive works characteristic of generative AI. A broad interpretation that includes AI training risks distorting the scope of the exception, effectively creating a compulsory license that the legislator never intended¹²⁹.

Both the definition in Article 2(2) and the explanatory Recital 8 of the DSM Directive restrict the scope of text and data mining to analytical techniques aimed at generating information. Generative AI, by contrast, absorbs and reassembles content to create synthetic outputs, exceeding the limits of what EU law classifies as lawful TDM. Generative AI requires large-scale, detailed reproduction and recombination of copyrighted content, unlike TDM methods, which generally access smaller datasets and focus on extracting factual or statistical information. The purpose of generative AI is to produce new outputs that resemble the training data, which is fundamentally different from the knowledge-generating aim of TDM and therefore falls outside the exception's scope. This distinction demonstrates the inapplicability of the TDM exception to AI training¹³⁰.

4.2.1.3. Opt-out mechanism problems

The opt-out mechanism established in Article 4(3) of the DSM Directive presents several systemic problems when assessed in light of the EU's policy objectives to promote both high-quality,

¹²⁸ See e.g. Paul Goldstein and others 'Kneschke vs. LAION - Landmark Ruling on TDM exceptions for AI Training Data - Part 2' (2024) Kluwer Copyright Blog <<https://legalblogs.wolterskluwer.com/copyright-blog/kneschke-vs-laion-landmark-ruling-on-tdm-exceptions-for-ai-training-data-part-2/>> accessed 30 July 2025.

¹²⁹ See e.g. Lucchi (n 14), 41; Tim W. Dornis, 'The Training of Generative AI is Not Text and Data Mining' (2024) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4993782>, 9 accessed 30 July 2025.

¹³⁰ Dornis, 'The Training of Generative AI' (n 129).

unbiased generative AI systems and fair remuneration for authors and rightsholders. First, this mechanism does not provide any guarantee of income increase to rightsholders: by opting out, rightsholders simply reserve their exclusive rights neutralising the TDM exception, which results in neither access for AI developers nor payment for creators – ‘a lose-lose scenario’¹³¹. This is likely to result in a shift towards licensing-based access to training content. However, the ability to obtain licenses for large volumes of copyrighted material is realistically available only to major AI developers with significant resources, potentially reinforcing existing market concentration and raising concerns about fair competition and access. Third, such structural bias in licensing access, combined with the large volume of opted-out content, risks reinforcing representational imbalances in AI training data. As a result, training is likely to rely on incomplete or non-representative datasets, increasing the risk that the output of AI systems will be discriminatory, biased, or otherwise distorted¹³².

4.2.1.4. From problem to policy: legislative recognition of the issues

As previously noted in Chapter III in June 2024, under the Hungarian Presidency, the Council of the European Union circulated a detailed policy questionnaire¹³³ to all Member States, seeking their official positions on the relationship between generative AI and copyright, including questions on text and data mining (TDM), licensing mechanisms, transparency, liability, and copyrightability of AI outputs.

A review of Member States’ positions, summarized in December 2024, reveals that significant legal and practical uncertainty exists around the application of the text and data mining (TDM) exception under Article 4 of the DSM Directive in the context of AI training. Several Member States questioned whether the current exception adequately covers AI training practices and noted that the TDM exception only applies to acts of reproduction and extraction, not to subsequent commercialization or public deployment of AI models. ‘In this context several Member States were of the view that copyright uses for AI training go beyond the scope of the TDM exception’.

Concerns were also raised about the vague definition of ‘lawful access’, particularly whether mere online availability qualifies as such, as well as about the practical functioning of the opt-out mechanism, which remains unclear. Furthermore, several Member States highlighted that the opt-out system has failed to create a meaningful licensing incentive and that the current legal

¹³¹ Martin Senftleben 'The TDM Opt-Out in the EU – Five Problems, One Solution' (2025) Kluwer Copyright Blog <<https://legalblogs.wolterskluwer.com/copyright-blog/the-tdm-opt-out-in-the-eu-five-problems-one-solution/>> accessed 30 July 2025.

¹³² *ibid.*

¹³³ Policy questionnaire (n 43).

framework does not allow for the effective removal of copyrighted content from already trained models. A number of Member States also emphasized the difficulties rightsholders face in monitoring whether their rights are being respected. Finally, doubts persist as to whether EU-based solutions can be effectively applied to AI systems trained outside the EU but deployed within the Union, despite Recital 106 of the AI Act affirming their applicability.

In April 2025, the European Parliament, through its Research Service and taking into account the results of the above-mentioned consultation with Member States, acknowledged ongoing concerns regarding the legal framework governing the use of copyright-protected materials in the training of general-purpose AI models. In its briefing ‘AI and copyright: The training of general-purpose AI’¹³⁴, the Parliament recognised that the current EU copyright regime, particularly the opt-out mechanism under Article 4 of the DSM Directive, fails to fully address the legal uncertainty surrounding text and data mining for AI training. The document reflects a growing awareness about the effectiveness of machine-readable opt-outs, the deployment of technological protection measures by rightsholders, which limit or prevent access for TDM purposes even when the use would otherwise be lawful. Furthermore, the document points out that the opt-out mechanism may be ineffective in cases where rights-holders do not control the website displaying the content and are therefore unable to implement the opt-out notice. Finally, it raises concerns about the duration of data retention, as GPAI developers may need to retain copies of copyrighted works not only for TDM—as intended by the DSM Directive—but also for subsequent stages such as model development.

Notably, the Parliament recognised the need for further legislative clarification and pointed to the planned 2026 review of the DSM Directive as an opportunity to address these shortcomings. Thus, the April 2025 resolution reflects the EU’s growing commitment to developing a more balanced and transparent legal framework for AI training practices involving protected works.

4.3. The use of copyrighted works in AI training under US law: the fair use debate

4.3.1. What is fair use

The fair use doctrine is a legal principle under US copyright law that permits the limited use of copyrighted material without the rightsholder's permission when such use serves socially beneficial purposes, such as criticism, comment, news reporting, teaching, scholarship, or

¹³⁴ European Parliamentary Research Service, ‘AI and copyright: The training of general-purpose AI’, PE 769.585, (2025) <[https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2025\)769585](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769585)> accessed 1 August 2025.

research. It is codified in Section 107 of the US Copyright Act¹³⁵ and evaluated on a case-by-case basis using four non-exclusive factors:

1. The purpose and character of the use, including whether it is commercial or for nonprofit educational purposes (while commercial uses are not presumptively unfair, they still tend to weigh against a finding of fair use¹³⁶), and whether the use is ‘transformative’.

A use is considered transformative if it adds new expression, meaning, or purpose to the original work, rather than replacing it. This concept was first formulated by the US Supreme Court in *Campbell v Acuff-Rose Music Inc*, where the Court held that a parody may qualify as fair use if it ‘adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message’¹³⁷. More recently, in *Andy Warhol Foundation for the Visual Arts Inc v Goldsmith*¹³⁸, the Court clarified that a use is not necessarily transformative merely because it conveys a different meaning or message; rather, the focus is on whether the new use serves a fundamentally different purpose.

2. The nature of the copyrighted work (e.g., factual vs. creative content).

Courts traditionally distinguish between factual or informational works and highly creative or expressive works. As established in *Harper & Row Publishers, Inc. v. Nation Enterprises*¹³⁹, the use of unpublished and highly creative material weighs more heavily against fair use, whereas the use of factual or published works is more likely to support a finding of fair use.

3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole.

The third fair use factor evaluates both the quantity and qualitative value of the portion used in relation to the copyrighted work as a whole. As the Supreme Court explained in *Harper & Row Publishers, Inc. v. Nation Enterprises*, even a small excerpt may weigh against fair use if it represents ‘the heart of the work’¹⁴⁰. Conversely, copying a large portion may be justified if it is necessary for the transformative purpose of the use. Courts assess whether the amount taken was

¹³⁵ 17 USC § 107.

¹³⁶ Lemley and Casey (n 21), 743, 765.

¹³⁷ *Campbell v Acuff-Rose Music Inc*, 510 US 569, 579 (1994).

¹³⁸ *Andy Warhol Foundation for the Visual Arts Inc v Goldsmith*, 598 U.S. 508 (2023).

¹³⁹ *Harper & Row Publishers Inc v Nation Enterprises*, 471 US 539 (1985) 545–549, 564.

¹⁴⁰ *ibid*, 564–566.

reasonable in light of the purpose of the secondary use, and whether the copied portion contained the most valuable or expressive elements of the original.

4. The effect of the use upon the potential market for or value of the original work.

The fourth fair use factor considers the effect of the secondary use on the potential market for, or the value of, the original work. If the secondary work may replace the original work or undermine existing or potential licensing markets, this factor weighs heavily against fair use. On the other hand, uses that do not target the same market or serve a different purpose are less likely to cause significant harm.

In the United States, opinions are divided as to whether the use of copyrighted works for training AI constitutes fair use. A wide range of arguments has been advanced both in favor of and against the application of the fair use doctrine to the training of generative AI systems.

4.3.2. Arguments for fair use

Supporters of applying the fair use doctrine argue that training AI models serves transformative and socially beneficial purposes¹⁴¹. They claim that use of the copyrighted materials for ML systems learning is transformative ‘because even though it doesn’t change the underlying work, it changes the purpose for which the work is used’¹⁴². This position centers on the idea that ML systems do not seek to exploit creative expression, but rather extract uncopyrightable elements such as facts, ideas, functional patterns, or linguistic structure¹⁴³. For example, language models learn grammar and sentence structure, not literary style; facial recognition systems learn to identify features, not to reproduce the composition or lighting of the exact photograph, or what the photographer was intending to depict in their photo¹⁴⁴. Thus, according to this view, AI training is not a replacement for the original works, but rather a means of extracting patterns and information that allow machines to generate novel content.

They further argue that this kind of learning is comparable to human learning, where reading or viewing copyrighted works to gain understanding or extract facts does not violate copyright¹⁴⁵.

¹⁴¹ See, e.g. Lemley and Casey (n 21); E. Lee, ‘Fair Use and the Origin of AI Training’ (2025), Houston Law Review, Santa Clara Univ. Legal Studies Research Paper No. 5253011, 123, <<https://ssrn.com/abstract=5253011>> accessed 22 August 2025.

¹⁴² Lemley and Casey (n 21), 748.

¹⁴³ Lemley and Casey (n 1), 772.

¹⁴⁴ *ibid.*

¹⁴⁵ See, e.g., Lemley and Casey (n 21) 751-753; US Copyright Office, Copyright and Artificial Intelligence Part 3: Generative AI Training (n 1), 42.

Supporters of fair use for AI training also emphasise that the application of this doctrine is necessary for advancing scientific research, innovation, and competitiveness¹⁴⁶. They also warn against attempts to impose licensing obligations on AI developers for the use of training materials. They argue that such a requirement would disproportionately benefit large corporations while excluding researchers, startups, and public-interest developers. The cost of negotiating licenses for billions of copyrighted materials would create significant barriers to entry into the market, potentially limiting the innovation opportunities and centralizing power in the hands of a few major players. They advocate instead for a fair use framework that supports openness, experimentation, and technological progress¹⁴⁷.

4.3.3. Arguments against fair use

On the other hand, critics of the fair use doctrine application argue that allowing unlicensed use of copyrighted works for AI training undermines the core principles of copyright law.

First of all, as it was discussed earlier in Chapter II, AI output could be similar to or even replicate the original works. In June 2025, a new lawsuit was filed by Disney and Universal against Midjourney in the U.S. District Court for the Central District of California, alleging extensive copyright infringement involving AI-generated images that closely resembled their iconic characters. The plaintiffs claim that ‘Midjourney, however, seeks to reap the rewards of Plaintiffs’ creative investment by selling an artificial intelligence (“AI”) image-generating service (“Image Service”) that functions as a virtual vending machine, generating endless unauthorized copies of Disney’s and Universal’s copyrighted works’¹⁴⁸.

Further, in many AI copyright disputes, critics highlight the commercial nature of AI model development and deployment, particularly emphasizing that these systems are designed to generate substantial revenue, which significantly weakens the fair use defence. For instance, in *Kadrey v. Meta Platforms, Inc.*¹⁴⁹, US District Court acknowledged that while training Meta’s Llama language model was highly transformative (Factor 1 favored fair use), the fact that the model was intended for commercial exploitation, generating potentially billions in revenue and enabling the creation of competing works, weighs strongly against a fair use finding¹⁵⁰. The Court warned that ‘no matter how transformative’ the use may be, ‘it’s hard to imagine that it can be fair use’ when

¹⁴⁶ See, e.g., Pamela Samuelson, ‘Does Training AI on Copyrighted Works Violate Copyright Law?’ (2024) 103 Boston University Law Review 1389, 1421.

¹⁴⁷ See, e.g., Lemley and Casey (n 21); US Copyright Office, Copyright and Artificial Intelligence Part 3 (n 5), 74, 75.

¹⁴⁸ *Disney Enterprises Inc v Midjourney Inc.*, 2:25-cv-05275, (C.D. Cal.) (Complaint), 2.

¹⁴⁹ *Kadrey v Meta Platforms Inc* (n 4).

¹⁵⁰ *ibid*, 18,19.

copyrighted materials train tools that flood the market with the content capable of displacing the original works’¹⁵¹.

The Association of American Publishers asserted in their comments provided to the U.S. Copyright Office that ‘Gen AI training . . . does not extract the ideas, facts, or concepts being conveyed by an author, it solely extracts the exact expressive choices made to convey those ideas—i.e., the words an author used, and the order in which they were placed’¹⁵².

4.3.4. The position of the U.S. Copyright Office

The U.S. Copyright Office analysed all four factors of the fair use doctrine concerning the use of copyrighted works for AI training to assess whether fair use could apply to such use and made the following preliminary conclusions in its pre-published report¹⁵³.

As regards the transformativeness, the U.S. Copyright Office suggests that it depends not merely on the act of copying, but on the ultimate purpose for which the copy is used. In the U.S. Copyright Office’s view, training an AI model on a large and diverse dataset may often be transformative, especially when the model is deployed for functions different from the original works’ purpose such as grammar correction, content moderation, or assisting language learners. However, training models to generate outputs that closely resemble or even replicate copyrighted works, such as generating images based on popular animated characters, is unlikely to qualify as transformative. ‘Because generative AI models may simultaneously serve transformative and nontransformative purposes, restrictions on their outputs can shape the assessment of the purpose and character of the use’¹⁵⁴.

The U.S. Copyright Office also rejects two prevailing arguments: that AI training is transformative because it is not for expressive purposes, and that it is similar to human learning. The Copyright Office suggests that language and image models are trained on huge quantities of expressive content, such as sentence structure, aesthetic composition, and linguistic nuance, with the clear purpose of generating new expressive outputs. Therefore, the use cannot be deemed ‘non-expressive’. Second, the U.S. Copyright Office rejects the analogy between AI training and human learning. It argues that fair use does not automatically permit extensive copying, even for educational purposes, and that computers should not receive broader exceptions than humans.

¹⁵¹ *ibid*, 39.

¹⁵² U.S. Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training (n 1), 44.

¹⁵³ *ibid*, 45-48.

¹⁵⁴ *ibid*, 46.

Moreover, AI training differs fundamentally from human learning, as it involves making perfect digital copies and processing them at superhuman speed and scale.

In assessing commercial use, the focus should not be on whether the organisation is for-profit or non-profit, but on the purpose of the training, model development, or even the specific output. For example, a commercial company may publish research without directly monetizing the model, making any commercial benefit indirect. Conversely, nonprofit entities may still engage in commercial activity if they license or sell access to models or datasets. Thus, the analysis must focus on the actual use and its function in practice, rather than on the formal classification of the organization conducting it.

As regards the amount and substantiality of the portion used, the U.S. Copyright Office said that while the use of complete works generally counts against fair use, it may be justified in the context of transformative, large-scale AI training where the copied content remains inaccessible.

Finally, as regards the fourth factor, the U.S. Copyright Office notes that AI training involving copyrighted works can cause significant market harm, especially when models generate outputs that either directly substitute for original works or dilute the market for stylistically similar content. The risk of harm is higher when licensing for training is feasible, since the availability of such licenses weighs against fair use under the fourth statutory factor.

Thus, after evaluating all four fair use factors, the U.S. Copyright Office reached the following conclusion: fair use must be assessed on a case-by-case basis, with particular emphasis likely to fall on the first and fourth factors. While some uses of copyrighted works in AI training may qualify as fair use, especially noncommercial research and analysis uses, others, such as unauthorized copying from pirate sources for commercial exploitation, are unlikely to meet the standard. Most cases will fall between these extremes¹⁵⁵. The U.S. Copyright Office suggests that the determination of fair use must be context-specific and sensitive to how copyrighted works are used at each stage of model development and deployment¹⁵⁶.

As noted above, the third report is a pre-publication version, and the final version has not yet been released. Just days after the U.S. Copyright Office released this preliminary version of the Report President Trump abruptly dismissed Shira Perlmutter, the Register of Copyrights, along with the

¹⁵⁵ U.S. Copyright Office, Copyright and Artificial Intelligence Part 3: Generative AI Training (n 1), 42, 74.

¹⁵⁶ *ibid*, 107.

Librarian of Congress, Carla Hayden. Multiple reports link these firings to disagreement with the report's conclusions¹⁵⁷.

4.3.5. Observations

At the July 2025 AI Summit, President Trump strongly opposed the idea of requiring AI companies to pay for every copyrighted work used in training datasets. He argued that such a requirement would be ‘not doable’ and would undermine the competitiveness of the US AI sector, especially in relation to countries like China that do not impose similar constraints. Trump emphasised that reading and learning from copyrighted materials should not trigger payment obligations or licensing negotiations. He also called for a unified federal regulatory framework for AI, warning that inconsistent or overly strict state-level rules could hold back innovation and hinder national progress in the field¹⁵⁸.

Therefore, it remains uncertain whether the Copyright Office’s position may change during the preparation of the final report. Nonetheless, it is desirable that the Copyright Office preserve its independent judgment in order to maintain a balanced approach that reflects the interests of both rights holders and technology developers.

4.4. Significant court cases: analysis of recent landmark court decisions

As legal uncertainty surrounding the use of copyright-protected materials in AI development intensifies, approximately 70 lawsuits are currently pending in U.S. courts concerning alleged copyright infringement in the training of AI models, including several class actions. A significant number of similar cases have also been filed or are already under examination in various European jurisdictions. As discussed above, legislators have so far been reluctant to introduce substantial regulatory changes, largely because the field remains insufficiently understood, and most governments prefer to await judicial clarification of the key legal questions. As a result, case law will play a decisive role in the near future in shaping how core issues, such as the copyrightability of AI-generated outputs and the permissibility of using copyrighted works for AI training, are ultimately resolved. For this reason, a comprehensive analysis of the current legal landscape requires close attention to ongoing litigation. Several cases where judgments have already been

¹⁵⁷ See, e.g., The Guardian, ‘Trump Fires Copyright Office Chief Shira Perlmuter’ (12 May 2025). <https://www.theguardian.com/us-news/2025/may/12/trump-fires-copyright-office-shira-perlmutter>>, accessed 05 August 2025; New York Post, ‘Trump Fires Head of US Copyright Office, Stoking Fears about Deregulation of AI’ (12 May 2025) <<https://nypost.com/2025/05/12/business/trump-fires-head-of-us-copyright-office-stoking-fears-about-deregulation-of-ai/>> accessed 05 August 2025.

¹⁵⁸ AI Summit 2025, ‘President Trump on Copyright and AI Regulation’ (YouTube, 23 July 2025) <<https://www.youtube.com/watch?v=BrTT7dX0mcQ>> accessed 05 August 2025, 20:09–25:46.

issued or are still pending will be examined below, as they are widely discussed and are beginning to influence the positions of legislators, legal scholars, and market participants. While these rulings are far from universally accepted and continue to provoke strong public debate and dissatisfaction among certain stakeholders, they represent the first concrete legal interpretations of the challenges posed by AI in the copyright domain. Thus, they offer valuable insights into how courts are beginning to approach issues of copyright infringement in the context of AI training and output.

1. *Kadrey v Meta Platforms Inc*¹⁵⁹ and *Bartz v Anthropic PBC*, *Bartz v. Anthropic PBC*, 3:24-cv-05417, (N.D. Cal. 2025).

In *Kadrey v Meta Platforms Inc*, thirteen published authors filed a lawsuit against Meta for copyright infringement, claiming that the company used their copyrighted books, obtained from ‘shadow libraries’ like Books3 and Anna’s Archive, to train its large language models (LLMs) known as Llama. Meta acknowledged the copying but invoked the fair use doctrine as a complete defence.

The court analyzed all four statutory fair use factors. On the first factor (the purpose and character of the use), the court found that Meta’s use was highly transformative: the books were not used for reading but as data to develop a multifunctional AI tool. Although the use was commercial, the court held that its transformative nature outweighed this consideration. The second factor (the nature of the copyrighted work) favored the plaintiffs, as the works were creative and expressive. However, the court noted that this factor carries less weight, particularly where the works have already been published. Regarding the third factor (the amount and substantiality of the portion used), the court found in favor of Meta, holding that copying entire works was reasonable and necessary for the transformative purpose of training a language model. On the fourth factor (the effect of the use on the potential market), the court concluded that the plaintiffs had failed to present meaningful evidence of market harm, including harm from potential market substitution, lost licensing opportunities, or market dilution.

Based on this analysis, the court granted summary judgment in favor of Meta. It emphasized, however, that this decision was limited to the claims of these specific plaintiffs and did not establish the general lawfulness of using copyrighted material to train AI models.

At the same time as the decision in *Kadrey v Meta Platforms Inc*, the U.S. District Court for the Northern District of California issued a ruling in another similar case *Bartz et al. v Anthropic PBC*.

¹⁵⁹ *Kadrey v Meta Platforms Inc* (n 4).

In this case, the plaintiffs, Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, filed a lawsuit for copyright infringement against the defendant, Anthropic PBC. The plaintiffs alleged that Anthropic copied their books without authorization from both pirated and lawfully purchased sources. Anthropic then compiled these copies into a so-called ‘central library’, from which it created various datasets used to train large LLMs. Notably, the company retained the entire library as a permanent, general-purpose resource, even after ceasing to use certain works for training. All of the copying and use occurred without any license or permission from the copyright holders.

In this case, the court held that the fair use doctrine could apply, as all statutory factors, except for the nature of the copyrighted work (factor two), favored fair use. The court also emphasised the highly transformative nature of LLM technology. It further concluded that converting lawfully purchased print books into digital copies likewise constituted fair use, as the digital versions were created for internal use only after the physical copies had been destroyed. In this context, factor one (purpose) and factor three (amount) supported fair use, factor two slightly weighed against it, and factor four (market effect) was considered neutral. However, the court explicitly rejected the application of fair use to Anthropic’s use of pirated copies, holding that such use was not justified under the doctrine.

As a result, summary judgment was granted in favor of Anthropic only with respect to the training uses and digitized purchased books. However, summary judgment was denied regarding the use of pirated copies, and this issue will proceed to trial, including a determination of actual or statutory damages.

Thus, although the courts in these cases recognised fair use in the context of training artificial intelligence models using the books, the decisions were not issued in a class action and apply only to the individual plaintiffs. In *Kadrey v Meta Platforms Inc* the court explicitly emphasised that this ruling does not preclude other rightsholders from bringing similar claims, and that if future plaintiffs succeed in demonstrating harm from potential market substitution or market dilution, such use may not qualify as fair use. Therefore, while these are the first judicial decisions addressing fair use in the context of AI training, they do not resolve the main question of the fair use doctrine’s applicability to AI training in general. Future cases may well lead to opposite conclusions.

2. *Thomson Reuters Enterprise Centre GmbH et al v. ROSS Intelligence Inc.*, No. 1:2020cv00613, 547 (D. Del. 2023).

In this case, the plaintiff, Thomson Reuters, the owner of the Westlaw legal research platform, sued Ross Intelligence Inc., a legal tech startup, for copyright infringement. Ross sought to develop an AI-powered legal search engine and, after being denied a license to Westlaw's content (as Westlaw considered Ross a competitor), acquired a dataset known as the 'Bulk Memo', which included approximately 25,000 judicial summaries ('headnotes') derived from Westlaw. Ross then used these headnotes to train its AI system. Thomson Reuters brought claims for direct copyright infringement, arguing that the Westlaw headnotes were original and protectable works.

The U.S. District Court for the District of Delaware conducted a full statutory fair use analysis and concluded that Ross's use of the headnotes was not protected under fair use. The court found that factor 1 (the purpose and character of the use), weighed against Ross, as the use was commercial and not sufficiently transformative; Ross used the content to develop a competing legal research tool that replicated Westlaw's function. Factor two (the nature of the copyrighted work), slightly favored Ross, as the headnotes, though factual in nature, involved some creative elements and were sufficiently original to be copyrightable, but 'not that creative'. Factor three (the amount and substantiality of the portion used) also favored Ross, since the headnotes were not publicly distributed but only used internally for AI training. However, factor four (the effect on the market), strongly favored Thomson Reuters, as Ross's product was intended as a direct substitute and posed a threat to both existing and potential licensing markets for Westlaw's editorial content.

Finally, the court found that the first and fourth statutory fair use factors outweighed the others and thus rejected Ross's fair use defence. Summary judgment was granted in part for Thomson Reuters on the issue of infringement with respect to headnotes. This decision marks one of the first federal rulings in the United States to reject a fair use defence in the context of AI training. It demonstrates that uses deemed commercially competitive and insufficiently transformative may not qualify for fair use, even when the underlying data is not publicly distributed.

However, it should be noted that, unlike the AI systems in *Kadrey v. Meta Platforms Inc.* and *Bartz v. Anthropic PBC*, which involved generative AI capable of producing new expressive content and were therefore deemed highly transformative, the AI system developed by Ross in *Thomson Reuters v. Ross Intelligence Inc.* was not generative in nature. This distinction may have worked against Ross, as the court found the use to be commercial and insufficiently transformative under the first fair use factor. Unlike generative systems, Ross's tool did not create novel outputs but rather sought to replicate the core functionality of Westlaw, thereby undermining its fair use defence. Therefore, the decision in this case may become a landmark for AI systems that produce specialised content, but it is not decisive for generative AI.

3. *Allen v Perlmutter*, No 1:24-cv-02665 (D. Colo, filed on 26 September 2024).

Courts have also been addressing the question of the copyrightability of AI-generated outputs. As discussed above, in *Thaler v Perlmutter*, the court held explicitly that a work created without any human authorship cannot be protected by copyright. However, beyond fully autonomous AI-generated works, a number of ongoing cases concern creative outputs developed with partial human involvement, where the U.S. Copyright Office refused to register the works. One such case is *Allen v Perlmutter*, in which the plaintiff challenges the Office’s denial of copyright registration for a comic book created with the assistance of AI tools, raising the question of whether and to what extent human–AI collaboration may qualify for protection under the US copyright law.

The plaintiff, artist Jason M. Allen, created an artwork titled *Théâtre D’opéra Spatial* using the generative AI platform Midjourney. Allen refined the AI output through over 600 prompt revisions and performed extensive post-processing in Adobe Photoshop and Gigapixel AI. In September 2022, he applied for copyright registration, identifying himself as the author. The U.S. Copyright Office denied the application, stating that the work contained too much AI-generated material and lacked the required human authorship. Allen sought reconsideration twice, but both the Office and its Review Board upheld the denial, concluding that significant AI-generated content must be excluded or disclaimed, which Allen refused to do.

In September 2024, Allen filed a federal lawsuit in the District of Colorado seeking a judgment that *Théâtre D’opéra Spatial* qualifies for statutory copyright protection. Allen argued that his creative inputs—prompt engineering, selection among AI outputs, and post-production—constitute sufficient human authorship.

The U.S. Copyright Office ‘acknowledges that the process of prompting can involve creativity’, they confirm that ‘some prompts may be sufficiently creative to be protected by copyright’ as literary works¹⁶⁰. However, the Office specifies that ‘that does not mean that providing text prompts to Midjourney “actually form[s]” the generated images’¹⁶¹.

Thus, although the U.S. Copyright Office acknowledges that works created with the assistance of artificial intelligence may be copyrightable if they include sufficient human authorship, it has in some cases refused to register such works when the human contribution is deemed insufficient.

¹⁶⁰ U.S. Copyright Office, Registration Decision Regarding “*Théâtre D’opéra Spatial*” by Jason M. Allen (Copyright Review Board, 5 September 2023) <<https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf>> accessed 07 August 2025.

¹⁶¹ *ibid.*, 7.

This highlights the unresolved and highly complex question of how the U.S. Copyright Office and ultimately the courts will evaluate the threshold and nature of human creative input in AI-assisted works. The pending case of *Allen v Perlmutter* may provide further clarity on this issue; however, no ruling has been issued to date, and its outcome remains uncertain.

4. *Like Company v. Google Ireland* (EU, pending before CJEU)¹⁶²

In this case, the applicant, a Hungarian news media publisher, filed a lawsuit against Google Ireland, operator of the generative AI chatbot Gemini, alleging that the chatbot unlawfully reproduced and made available press publications protected by copyright and related rights. The dispute centers on summaries of articles generated by Gemini in response to user queries, which allegedly reproduced substantial parts of the applicant's content without consent or remuneration.

The claimant argues that such summaries exceed the permissible use of 'individual words or very short extracts' under Article 15 of the DSM Directive and therefore require prior authorisation and compensation. Additionally, the claimant argues that the chatbot's replies constitute an act of communication to the public, even when those responses appear not on a standard search engine results page but within the chatbot's own conversational interface, where they display protected content.

Google Ireland, in turn, denies any infringement, arguing that the chatbot does not retrieve or store protected content directly, but operates as a generative tool that transforms data into tokenized formats. It contends that its responses do not constitute reproduction or communication to a 'new public' and that the output falls within exceptions such as temporary acts of reproduction (Article 5(1) of InfoSoc Directive) and text and data mining under Article 4 of the DSM Directive. Google also maintains that much of the output is 'hallucinated' content not derived from the protected publications.

The referring Hungarian court submitted a request for a preliminary ruling to the Court of Justice of the European Union, seeking clarification whether the display of excerpts from press publications by an LLM-based chatbot constitutes a communication to the public within the meaning of Article 15(1) of the DSM Directive and Article 3(2) of the InfoSoc Directive and whether the training of such chatbots constitutes an act of reproduction under Article 2 of InfoSoc

¹⁶² C-250/25 *Like Company v Google Ireland* (Request for a preliminary ruling, pending, 3 April 2025) <<https://curia.europa.eu/juris/showPdf.jsf?docid=300681&doclang=EN>> accessed 07 August 2025; Request for a preliminary ruling from the Budapest Környéki Törvényszék (Hungary) lodged on 3 April 2025 – *Like Company v Google Ireland Limited* (Case C-250/25, *Like Company*) (C/2025/3039) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62025CN0250>>, accessed 07 August 2025.

Directive, and if so, whether this use qualifies for the text and data mining exception under Article 4 of the DSM Directive. Finally, the court seeks to determine whether the generation of chatbot responses that reproduce content from press publications, when prompted by user input, should also be regarded as reproduction for which the chatbot provider may be liable.

This is the first case in which a Member State court has referred questions to the CJEU concerning EU copyright law in the context of generative AI, particularly with respect to the applicability of the text and data mining exceptions to AI training. The outcome of this decision may determine how generative AI systems can lawfully interact with copyright-protected content in the EU, especially whether the training and output phases require licenses or fall within the scope of statutory exceptions.

5. *Rozsudek Městského soudu v Praze sp*¹⁶³

A plaintiff (anonymous law firm) sought judicial recognition of authorship and exclusive rights to an image generated by the DALL-E AI system, based on a textual prompt. The image was later used without consent by the defendant. In a 2024 ruling, the Municipal Court in Prague (*Rozsudek Městského soudu v Praze, sp. zn. 10 C 13/2023-16*) rejected a claim. The court held that the claimant failed to provide sufficient evidence proving authorship of the AI-generated work, particularly the exact input used and the causal link between the prompt and the resulting image. Crucially, the court found that a work created entirely by AI, without human creative expression, cannot qualify as a copyrightable work under Czech law. It emphasized that only natural persons can be considered authors, and the claimant's prompt was regarded as an unprotectable idea rather than a creative work. Consequently, the court dismissed all related claims, including demands for injunctive relief and removal of the image from the defendant's website.

This is the first court decision in the EU directly addressing AI-generated content and copyright. It confirmed that purely AI-generated works are unprotectable, and that human prompts alone do not establish authorship without significant creative input. While the court did not categorically rule out protectable AI-assisted works, it clearly requires concrete proof of human-directed creativity—beyond simple prompts.

6. Conclusion

¹⁶³ *Rozsudek Městského soudu v Praze sp zn 10 C 13/2023–16* (Czech Republic, 11 October 2023) (Municipal Court, Prague – Judgment No. 10 C 13/2023-16 (11 Oct 2023)) https://msp.gov.cz/documents/14569/1865919/10C_13_2023_10/108cad3e-d9e8-454f-bfac-d58e1253c83a accessed 08 August 2025.

Thus, at present, judicial practice is in the process of emerging in both Europe and the United States, and it is likely to provide answers to questions concerning the copyrightability of AI-generated content and the legality of using copyright-protected works for training artificial intelligence systems. In the United States, the first decisions have already been issued under the doctrine of fair use. However, these rulings do not qualify as landmark precedents, as the claims were not brought as class actions and the decisions affect only the specific plaintiffs involved. Moreover, dismissals in these cases were based on highly case-specific grounds. For instance, in *Kadrey v Meta Platforms Inc*, the court expressly noted that in this particular case the plaintiffs had failed to provide sufficient evidence of market harm, which does not preclude plaintiffs in other cases from successfully proving the contrary.

Equally significant are cases addressing the boundaries of copyrightability where the output involves partial human involvement. These proceedings raise the question of what level of human contribution will be considered sufficient to provide copyright protection for such works. The outcomes of these cases remain pending, but the decisions could define the future development of both judicial practice and legislative frameworks in this area.

4.5. Approaches taken by other market participants, such as collective rights management societies

Artists, composers, visual creators, and other professionals from the creative sector have been actively engaged in the debate surrounding the authorship of AI-generated content and the use of copyrighted works for AI training. Their interests are generally represented by various collective rights management organisations and professional associations. On the question of authorship, these bodies unanimously support the official legal and court's positions that only a human can be considered the author of a copyrighted work. Nevertheless, one exception attracted public attention: in 2017, the French collective rights society SACEM registered musical works generated by the AI system AIVA (Artificial Intelligence Virtual Artist)¹⁶⁴, formally attributing authorship to the AI system. Although this case was often cited as a precedent for recognising AI as an author, the registration was in fact made through a human representative and did not reflect a shift in legal doctrine. As for today, SACEM's current official position, published on its website, makes clear that music and lyrics 'generated entirely by an artificial intelligence tool is not protected by copyright and cannot be registered with Sacem'¹⁶⁵.

¹⁶⁴ AI Music Outputs (n 71), 52, 53.

¹⁶⁵ SACEM, SACEM and Artificial Intelligence (May 2025) <<https://www.createursediteurs.sacem.fr/en/news/our-society/sacem-and-artificial-intelligence>>, accessed 6 August 2025.

With regard to the training of artificial intelligence systems, collective management organisations and professional associations, both in the United States and in the EU, have been actively opposing the use of copyright-protected works without the consent of rightsholders and fair compensation. These organisations argue that such practices undermine the fundamental principles of copyright law and jeopardise the economic interests of creators whose works are being exploited on a massive scale by AI developers without transparency or remuneration.

For example, the European Writers' Council (EWC), together with other creative organisations (such as ECSA (European Composer and Songwriter Alliance, EFJ (European Federation of Journalists, FERA (Federation of European Screen Directors) and others), has strongly criticised the current application of the text and data mining (TDM) exception under Article 4 of the DSM Directive, particularly its use by developers of generative AI systems. In their Joint letter to the European Parliament's JURI Committee 'Protecting the rights of creators and artists in the development of generative AI'¹⁶⁶, the EWC argued that the opt-out mechanism applied to TDM enables mass exploitation of protected works without prior consent, leaving authors with no real control over the use of their intellectual property. As the Council emphasised, 'we have never experienced copyright exceptions so unclear, so widely misused and so damaging to our creative communities. This is not just unfair – it is unacceptable'¹⁶⁷. This interpretation, they argue, has a devastating impact not only on authors and performers but also on AI developers, who face legal uncertainty and potential liability due to the lack of transparent licensing frameworks. More broadly, the current regime allows generative AI models 'to put the cart before the horse' by relying on copyright exceptions without giving creators the opportunity to exercise their rights. In this context, the EWC has highlighted several unresolved legal questions, including the scope of TDM uses permitted in relation to generative AI, the identity of rightsholders entitled to opt out, and whether rights reservations can be enforced retroactively. The Council calls for a legislative shift from the current opt-out system to an opt-in model that would ensure that authors and performers can actively authorise the use of their works in AI training. It also advocates for the development of functional and proportionate remuneration mechanisms to guarantee fair compensation on the market.

Another example is an issue by a broad coalition of European creators, composers, performers, authors, publishers, producers, and other rights-holders' organisations a 'Joint statement regarding

¹⁶⁶ European Writers' Council, 'Joint letter to the European Parliament's JURI Committee on copyright and generative artificial intelligence – opportunities and challenges' (19 June 2025) <https://europeanwriterscouncil.eu/2506_jl_juritoep/?utm_source> accessed 8 August 2025.

¹⁶⁷ *ibid.*

the AI Act implementation measures adopted by the European Commission'¹⁶⁸ published on 30 July 2025 and strongly criticising the European Commission's implementation measures under the AI Act, including the GPAI Code of Practice, Guidelines, and training-data summary Template. They argue that despite extensive engagement with right-holders communities, the final regulatory package fails to comply with the core objectives of Article 53(1)(c) and (d), which were intended to empower creators to exercise and enforce their rights in response to widespread unlicensed use of protected content. Rather than delivering meaningful protection, the implementation measures are seen as benefiting generative AI providers at the expense of creators, undermining the AI Act's promise to balance creativity with innovation. The coalition called on the Commission to revise the package to ensure effective enforcement of the AI Act, and to center future regulation on the principles of authorisation (opt-in), transparency, and fair remuneration for authors and performers.

In US the positions of creative industry representatives are similar. For instance, the US Authors Guild opposes the unlicensed use of copyrighted works for training generative AI systems¹⁶⁹. The organisation argues that incorporating books and articles into AI technologies without the author's consent, compensation, or credit is unfair and harmful to the literary profession. It warns of the risk of market saturation with mass-produced AI-generated texts that could devalue human-authored work and threaten the writing profession. To address these concerns, the Guild advocates for a legal and regulatory framework grounded in five core principles: consent, compensation, transparency, protection of identity in AI outputs, and labelling of AI-generated content. Specifically, the Guild proposes the establishment of a collective licensing regime whereby authors who opt-in would be represented by a collective management organisation (CMO), which would negotiate with AI companies and distribute royalties for both past and future uses of their works. The Guild opposes any form of copyright or sui generis protection for AI-generated outputs, arguing that this would further undermine human creators.

A notable licensing proposal was made by the German collective management organisation GEMA in 2024. GEMA introduced a two-component licensing model for AI services, created to ensure fair compensation for the use of music in both the training and exploitation phases of AI systems. Under the first component, AI providers are required to obtain a licence covering the use of protected musical works in training datasets, regardless of the technical format or the location

¹⁶⁸ Coalition of European creator organisations, 'Joint statement regarding the AI Act implementation measures adopted by the European Commission' (Brussels, 30 July 2025) <<https://composeralliance.org/news/2025/7/joint-statement-regarding-the-ai-act-implementation-measures-adopted-by-the-european-commission/>> accessed 6 August 2025.

¹⁶⁹ Authors Guild, FAQs on the Authors Guild's Positions and Advocacy Around Generative AI (2025) <<https://authorsguild.org/advocacy/artificial-intelligence/faq/>> accessed 6 August 2025.

of the training process. The model establishes that rightsholders are entitled to 30% of all net income generated by the AI system, including revenues from subscriptions or any other form of commercial exploitation of the AI-generated output. To ensure a minimum level of fairness, GEMA's model includes a baseline royalty that applies even when the AI system generates little or no profit. This minimum payment is calculated based on the amount of output the AI system produces, ensuring that rightsholders are compensated proportionally to the system's use of their works. The second component of the model covers the subsequent commercial use of AI-generated content, such as background music or monetised releases. GEMA's approach confirms that authors must receive a share of income from such uses that is at least equal to the royalty they would have earned for comparable human-generated works.

Thus, the positions expressed by authors' associations and collective management organisations clearly reflect the growing concern within the creative industries about the scale at which artificial intelligence is being integrated into areas of cultural production that have traditionally relied exclusively on human creativity. On the one hand, these stakeholders consistently emphasise their openness to and support for technological innovation. On the other hand, they identify a profound threat to the sustainability of the entire creative ecosystem and the livelihoods of millions of its participants. In this context, their insistence on the need for appropriate remuneration and transparency when copyrighted works are used for AI training is both understandable and justified, particularly given that such training would not be possible without the intellectual contributions of authors and, in fact, carried out at their expense.

4.6. Conclusion: Comparative reflections on copyright, AI-generated content, and training uses in the EU and the US

The integration of generative artificial intelligence into creative and technological domains presents profound legal challenges for both the European Union and the United States. While their copyright systems differ in structure and administration, both jurisdictions have adopted a human-centred approach to the copyrightability of AI-generated content and a cautious approach to the possibility of using protected works for AI training.

Both EU and US frameworks share the same core principle: only human beings can be recognised as authors under existing copyright law. The Court of Justice of the European Union has consistently interpreted originality as the author's own intellectual creation, which includes personality, free and creative choices, and human expression. Similarly, in the US, courts and the U.S. Copyright Office have reaffirmed that the statutory framework requires human authorship.

AI-generated content, produced without meaningful human creative input, does not qualify for protection in either system and generally falls into the public domain.

At the same time, both jurisdictions acknowledge that AI-assisted outputs may be eligible for protection, but only to the extent that human creativity is objectively reflected in the final result. In both systems, authorship claims must be assessed case-by-case, and neither has developed a consistent standard for evaluating creativity in prompt engineering or post-editing.

This approach has significant implications: it creates legal uncertainty, particularly in borderline cases where human input is minimal or indirect. As scholars such as Lemley have observed, if copyright protection hinges on the creativity of a prompt rather than on the output itself, authorship becomes difficult to determine, and the boundaries of protection become practically indistinguishable.

Another key legal challenge examined in this Chapter concerns whether and how copyright-protected works may be lawfully used to train generative AI systems. Here, the difference between the EU and US systems becomes more evident.

In the EU, Article 4 of the DSM Directive, together with the AI Act, establishes a TDM exception with an opt-out mechanism for rights holders. While this allows the usage of copyrighted materials for AI training, it has proven to be both technically and conceptually unclear. The lack of a clear definition of ‘machine-readable’ opt-outs, the uncertainty around lawful access, and licensing could undermine the effectiveness of the regime. Furthermore, as multiple Member States pointed out, the TDM exception was never designed for generative AI and does not adequately address issues of reuse, commercialisation, or exploitation of the trained models. There is now a growing consensus that legislative reform will be required to adapt the copyright framework to the realities of large-scale AI training.

In the US, no specific statutory provision governs the use of protected works for AI training. Instead, the permissibility of such uses is analysed under the flexible doctrine of fair use. Recent court decisions (e.g., *Kadrey v Meta*, *Bartz v Anthropic*, *Thomson Reuters v Ross Intelligence*) have demonstrated that AI training may, under certain circumstances, qualify as fair use, particularly when the use is deemed transformative, the works are not reproduced in the outputs, and market harm is unproven. However, where training uses are commercial, competitive, or based on pirated materials, courts have shown an intention to reject the fair use defence. These cases underline the fact that fair use remains an uncertain and contested defence in the context of AI training, heavily dependent on facts, purpose, and the scale of use.

Beyond courts and lawmakers, collective rights management organisations and creative industry associations have emerged as powerful voices in this debate. In both the EU and the US, they have rejected the legitimacy of unlicensed AI training and demand frameworks based on consent, transparency, and fair remuneration. Organisations such as GEMA (Germany), SACEM (France), and the Authors Guild (US) have proposed licensing models or regulatory reforms that would enable authors to monetise the use of their works in AI training, emphasising the injustice of exploiting human creativity to train machines without compensation.

At the same time, attention is focused on the courts, where high-profile cases on these issues are in process. While early rulings offer some clarity, they also raise concerns about inconsistency, jurisdictional fragmentation, and the increasing burden on courts to evaluate complex technological processes using probably outdated legal concepts.

The analysis in this Chapter suggests that while copyright law still serves its main purpose of protecting human creativity, it is not adapted well to address the new challenges related to AI. In both jurisdictions, there is a growing recognition that a new legal category may be required, whether as a *sui generis* right for AI-generated outputs or a specialised licensing regime for training data. Without such reform, both creators and AI developers will be left without adequate legal certainty.

The future approach to addressing these challenges should balance encouraging innovation with protecting creators' rights. To remain relevant in the age of generative AI, copyright law needs to adapt to the interaction between humans and machines and create new ways to provide incentives and protection for new types of content and technological development.

V. Harmonisation

Generative AI has emerged as a technological breakthrough that will most likely have a profound impact on society and the global economy, possibly even greater than the advent of the internet. It is now developed primarily by the world's largest companies that make huge investments with the clear aim of deploying it globally rather than limiting it to a single domestic market. In an interconnected digital environment, where training materials for AI are sourced from the internet without regard to national borders, and where the quality of advanced AI systems depends on access to the broadest possible datasets to minimise bias, narrowness, and inaccuracies, it is important to ensure that AI can be developed without numerous obstacles because of the legislative differences and contradictions in all countries worldwide. Thus, it seems appropriate that rules governing AI training, as well as the allocation of rights in its outputs, should be aligned not only within individual regions, such as the European Union, but also internationally. Without such coordination, both rightsholders and AI developers face legal uncertainty, fragmented enforcement, and barriers to innovation, while harmonised standards would ensure fairness, legal clarity, and global interoperability in the use and governance of generative AI.

In the European Union, discussions address both integration within the Union itself and international cooperation. As noted in the European Parliament Resolution of 20 October 2020¹⁷⁰, Article 118 TFEU mandates the creation of measures to ensure uniform protection of intellectual property rights across the Union. The Resolution underlines that the EU level is the appropriate forum for regulating AI technologies to prevent fragmentation of the single market and differences in national rules. It emphasises that a fully harmonised EU regulatory framework could serve as a global legislative benchmark, and that new common rules for AI should take the form of a regulation to guarantee equal standards. The Parliament points out that legal certainty fosters innovation, while public trust in AI systems is essential to the sector's growth. Therefore, the regulatory framework should inspire confidence in the safety and reliability of AI while striking a balance between protecting the public and supporting business investment in innovation.

As a result, the AI Act was adopted in the form of a regulation rather than a directive. However, as discussed above, copyright legislation has not undergone substantive changes, except for extending the TDM provisions to the training of artificial intelligence. Copyright law in the European Union remains only partially harmonised, although significant progress in harmonisation has already been achieved. It appears appropriate that any future amendments to

¹⁷⁰ European Parliament, Resolution of 20 October 2020 (n 17).

copyright legislation concerning artificial intelligence should also be included in the broader plan for the further harmonisation of copyright provisions in the EU. This aligns with the objectives and directions set out in recital 31 of the InfoSoc Directive, which emphasises the need to maintain a fair balance of rights and interests between different categories of rightsholders, as well as between rightsholders and users of protected subject matter. The Recital notes that existing differences in exceptions and limitations across Member States can have direct negative effects on the functioning of the internal market for copyright and related rights, especially in the context of cross-border exploitation of works. To address this, it calls for such exceptions and limitations to be defined more harmoniously, with the degree of harmonisation determined by their impact on the smooth functioning of the internal market.

At the same time, in the United States, there is no harmonisation of legislation on artificial intelligence at the federal level. Most regulatory initiatives are developed and implemented at the state level, including rules on the ownership of AI-generated outputs, as seen in Arkansas. This fragmented approach risks creating uncertainty for developers and users, given the potential for divergent state requirements on training processes and rights allocation. At the federal level, authorities appear to be waiting for judicial developments before adopting comprehensive regulation; however, case law is evolving slowly, with decisions so far affecting only the parties involved and providing little guidance for broader legislative or judicial trends.

In such conditions, international cooperation, potentially under the coordination of organisations such as the World Intellectual Property Organization (WIPO), could promote unification in definitions, exceptions, and enforcement mechanisms. A consistent international framework would reduce fragmentation, facilitate cross-border licensing, and provide a more predictable legal environment for AI developers, rightsholders, and intermediaries. Such harmonisation would also help prevent ‘forum shopping’, where stakeholders seek jurisdictions with the most favorable rules, potentially undermining the integrity of copyright protection.

The idea of global harmonisation is not unrealistic. A historical precedent can be found in the Berne Convention, which established fundamental principles of copyright protection, including the criteria for copyrightability and the core standards of protection. Since its adoption in 1886, the Convention has been ratified by 181 countries, which have incorporated its basic provisions into their national laws, creating a broadly consistent international framework for copyright. Thus, the harmonisation of principles governing the training of artificial intelligence systems and the ownership of AI-generated outputs could follow a model similar to that used in the Berne Convention. Such an approach would establish minimum international standards, while allowing states to adapt specific mechanisms to their legal traditions.

Any such framework should be based on core principles: technology-neutrality, so that rules apply across different AI technologies; transparency, enabling traceability of training data and outputs; fair remuneration for both creative and investment input, and the familiar Berne principles of national treatment, automatic protection, and minimum standards.

A harmonised framework could begin with a lawful training exception for text and data mining, applicable to lawfully accessible works and subject to a uniform opt-out mechanism implemented through machine-readable means. Works excluded in this way should not be used for training without authorisation, and compliance with such signals would provide a defence against infringement claims in all participating jurisdictions.

Regarding AI-generated outputs, protection would depend on the presence of human authorship. Copyright would be granted only where a human exercises full creative control over the generation process and further processing and elaborating of AI-generated output, such that the output reflects the author's own intellectual creation. Outputs generated without such human involvement or minimal human involvement would either fall into the public domain or benefit from a *sui generis* neighboring right of limited duration, without moral rights and with a narrow scope to recover the investment. Comparative law already provides such analogies. For example, several Member States in the EU protect non-original photographs as related rights¹⁷¹; under the Database Directive, a special *sui generis* right is granted to database makers where the database results from a 'substantial investment' in obtaining, verifying, or presenting its contents, but not for creative input. In the UK, statute assigns authorship of computer-generated works to 'the person by whom the arrangements necessary for the creation of the work are undertaken'¹⁷². Thus, such specialised forms of protection for specific categories of intellectual property have proven effective and have been integrated into both legislation and practical application.

Additionally, a harmonised framework should ensure that copyright-protected works used for AI training are not exploited without fair remuneration to the right holders. Accordingly, the licensing mechanism could be established, allowing the use of such works in exchange for remuneration. This mechanism could operate through voluntary (negotiated and concluded by independent stakeholders) or collective licensing schemes, enabling rights holders to receive payments via collective management organisations¹⁷³ or other large datasets holders. As it is noted in U.S. Copyright Office report 'Copyright and Artificial Intelligence, Part 3: Generative AI Training'¹⁷⁴

¹⁷¹ European Commission, Trends and Developments in Artificial Intelligence (n 31), 92

¹⁷² Copyright, Designs and Patents Act 1988 (UK), UK Public General Acts 1988 c. 48, 9(3).

¹⁷³ U.S. Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training (n 1), 85.

¹⁷⁴ *ibid*, 103.

in recent years, a range of voluntary direct and collective licensing agreements for the use of copyrighted works in AI training has emerged. Some AI systems are now trained exclusively on licensed materials or works in the public domain¹⁷⁵. Such an approach would facilitate large-scale lawful access to training materials, reduce transaction costs, ensure high-quality materials for training, and guarantee that authors and other rights holders receive an equitable share of the revenue generated from AI systems trained on their works.

Another question that can and should be resolved is that the developers should be legally required to implement proportionate measures that prevent outputs from reproducing protected training content (verbatim or near-verbatim). Current policy instruments point in the same direction. The EU's new General-Purpose AI Code of Practice commits signatories to adopt 'appropriate and proportionate technical safeguards to prevent their models from generating outputs that reproduce training content protected by Union law on copyright and related rights in an infringing manner'¹⁷⁶.

The institutional architecture for such harmonisation could involve a WIPO-administered treaty, functioning either as a protocol to the Berne Convention or as a stand-alone agreement. This treaty would set the core definitions, principles, and minimum standards, accompanied by model laws for federal systems to ensure internal consistency.

The success of such a harmonisation effort would depend on balancing competing interests: preserving openness in training while allowing remuneration for high-value training materials; maintaining full copyright protection only for human-created works, but granting *sui generis* rights for non-copyright-protectable works to justify investments and stimulate technological development; and imposing transparency obligations that do not disproportionately burden small enterprises or research initiatives.

Nevertheless, it is evident that, at present, these measures are still at the discussion stage, with each attracting both supporters and opponents. However, the least controversial measures should start being implemented, as the European Union is doing now. At the same time, international cooperation is highly desirable to maintain consistency and avoid situations where some countries permit violations of authors' rights from other countries, which are protected under well-established global frameworks, while others create compelling barriers for developers, thereby hindering technological progress.

¹⁷⁵ *ibid.*

¹⁷⁶ Generative AI: Code of Practice (n 36), measure 1.4.

VI. Conclusion and Key Findings

This thesis sets out to examine the interaction between copyright law and artificial intelligence, focusing on five main research questions: (1) which existing legislation in EU and US applies to copyright related matters related to AI (2) whether and under what conditions AI-generated works can be protected by copyright; (3) how copyright frameworks in the EU and the US address the use of protected works for AI training; (4) what positions have been taken by public authorities, courts, and market stakeholders on these issues (5) is national or international harmonisation necessary to effectively address the challenges of interaction of copyright and AI to balance the interests of creative industry and technological development. The analysis combined legislative review, policy documents, and case law from both jurisdictions, providing a comparative perspective.

The analysis shows that, as of mid-2025, the European Union and the United States apply fundamentally different approaches to regulating the interaction between copyright and artificial intelligence.

In the EU, the legislator has adopted a harmonised regulation, the AI Act, ensuring a uniform approach to AI across all Member States. This regulation also addresses AI training by reinforcing the existing DSM Directive's text and data mining provisions and introducing enforceable transparency duties, respect for opt-outs, and documentation requirements for providers of general-purpose AI models. This reflects the EU's preference for a harmonised system, ensuring consistent obligations across all Member States, while still taking a cautious approach to broader legislative reform. The EU's strategy seeks to balance rightsholder protection with technological progress, although the suitability of applying TDM exceptions to AI training remains debated.

In the US, there is still no unified federal legislation specifically addressing copyright in the AI context. Instead, the legal environment is shaped by sector-specific statutes, executive orders, state-level initiatives, and the interpretative guidance of the U.S. Copyright Office. States such as Arkansas and California have introduced their own rules on ownership and transparency, creating the risk of fragmented, potentially conflicting obligations for developers operating across jurisdictions. The U.S. Copyright Office plays a central role in shaping policy and setting interpretative baselines, but its guidance is not binding on courts, leaving significant uncertainty until federal jurisprudence develops.

Overall, the EU approach is centralised and pre-emptive, embedding copyright safeguards within a uniform legislative framework, while the US approach remains decentralised and relies on administrative interpretation and litigation to define substantive rules.

As regards the question of AI-generated content legal status, both the EU and US frameworks share the same foundational principle: copyright protection requires human authorship. In the EU, the CJEU has consistently interpreted ‘author’ as a human creator who exercises free and creative choices. The U.S. Copyright Office and US courts likewise reject recognition of copyright for works generated without human involvement. There is, therefore, no legal or judicial uncertainty on this point.

Greater complexity arises when AI outputs involve some level of human creative input, such as detailed and original prompts or substantial post-generation modification. Where the human contribution meets the threshold of originality, the resulting work can be treated analogously to a derivative work and protected accordingly. However, when an AI output is used in its original form (as generated by AI), or the human modifications are minimal, the legal status is uncertain. Linking copyright protection to the creative level of a prompt presents significant practical and evidentiary challenges, with no clear or reliable criteria for assessment. In such cases, decisions would likely fall to courts or, in the US, to the U.S. Copyright Office at the registration stage, where both scenarios carry the risk of increased litigation and administrative burden.

At the same time, binding to the human authorship requirement does not resolve questions about the ownership of AI-produced content that does not qualify for copyright protection. Such works may still demand significant investment of time, effort, and resources, even if not of a creative nature. This raises the possibility that AI-generated outputs could be recognised as a distinct category of civil rights objects, warranting a separate legal regime connected to their specific characteristics.

In sum, there is no immediate need to reform the core principles of copyright law, which remain effective in protecting human authors and their creative works. However, the growing dissemination of AI-generated content without meaningful human involvement may require the development of a parallel framework to address the economic and legal realities of such outputs while preserving the human-centred foundation of copyright.

As regards the usage of the copyrighted materials for AI training, the comparative analysis reveals a clear difference between the EU and US approaches on this matter.

In the EU, the relevant framework is built on Article 4 of the DSM Directive, reinforced by the AI Act, which together establish a text and data mining exception combined with an opt-out mechanism for rightsholders. While this framework permits the use of protected works for AI training, it suffers from both technical and conceptual uncertainties, especially regarding the definition of ‘machine-readable’ opt-outs, the conditions for lawful access, and the role of licensing, which create practical enforcement challenges. Moreover, the TDM exception was never designed for generative AI purposes and fails to adequately address the issues of commercialization and exploitation of trained models. As acknowledged by several Member States, legislative reform will likely be necessary to adapt the copyright framework to the specific demands of large-scale AI training.

In the US, there is no statutory provision specifically addressing AI training. Instead, the permissibility of such uses is assessed under the doctrine of fair use. Recent litigation demonstrates that AI training may qualify as fair use where the use is transformative, does not result in the reproduction of the protected works in the outputs, and does not cause a proven market harm. Conversely, courts have signalled a readiness to reject the fair use defence when the training involves commercial or competitive uses, or relies on unlawfully sourced materials. As a result, fair use in this context remains highly fact-sensitive, unpredictable, and subject to ongoing judicial development.

Overall, the EU pursues a codified, rights-holder-oriented approach that seeks to harmonise obligations across Member States but faces significant questions of scope and effectiveness, whereas the US relies on a case-by-case judicial assessment under fair use, offering flexibility but generating legal uncertainty until more settled precedent emerges.

These differences in legal approach are also reflected in the positions of key market participants. Collective management organisations, professional associations, and author groups in both the EU and US support the principle that only humans can be recognised as authors of copyrighted works, rejecting any form of copyright or sui generis protection for AI-generated content. On the question of AI training, creative sector stakeholders are united in opposing the unlicensed use of protected works without consent, transparency, and fair remuneration.

Overall, the analysis suggests that the development of generative AI, and its cross-border nature, require harmonised legal standards both within individual jurisdictions and at the international level. Without such coordination, rightsholders and AI developers will face ongoing uncertainty, fragmented enforcement, and barriers to innovation. A harmonised approach could help ensure

fairness, legal clarity, and consistency across markets, while still allowing flexibility for technological progress.

Ultimately, the study shows that neither the centralised, pre-emptive model of the EU nor the decentralised, litigation-driven model of the US provides a fully adequate response on its own. An effective approach will depend on reaching a sustainable balance: protecting the legitimate interests of the creative sector, ensuring transparency and fair remuneration for the use of protected works, and maintaining an environment in which technological innovation can develop. Achieving this balance will require both national reforms and international dialogue, ideally leading to a framework that reflects shared principles while respecting different legal traditions.

Table of legislation, cases and other sources

1) Table of legislation

International Legislation

Agreement on Trade-Related Aspects of Intellectual Property Rights (15 April 1994) 1869, UNTS 299

Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, as amended 28 September 1979) 828 UNTS 221

WIPO Copyright Treaty (adopted 20 December 1996, entered into force 6 March 2002) 2186 UNTS 121

EU Legislation

Copyright, Designs and Patents Act 1988, UK Public General Acts 1988 c. 48

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20

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 [2001], OJ L 167/10

Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights [2006] OJ L372/12

Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs [2009] OJ L111/16

Directive 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 [2019] OJ 2019 L 130

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) [2024] OJ L1689/1

US Legislation

Arkansas, House Bill 1876, 95th General Assembly, *An Act Regarding the Ownership of Model Training and Content Generated by a Generative Artificial Intelligence Tool*, engrossed version (4 July 2025)

<<https://arkleg.state.ar.us/Home/FTPDocument?path=%2FBills%2F2025R%2FPublic%2FHFB1876.pdf>> accessed 29 July 2025

California, *Assembly Bill 2013, Generative artificial intelligence: training data transparency* AB-2013,
<https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013>
accessed 30 July 2025

California, *Assembly Bill 412 (2025-2026 Regular Session), Generative Artificial Intelligence: Training Data: Copyrighted Materials* AB-412 (amended 7 May 2025)
<https://leginfo.legislature.ca.gov/faces/billHistoryClient.xhtml?bill_id=202520260AB412>
accessed 30 July 2025

California, *Senate Bill 942 California AI Transparency Act (2023-2024 Regular Session)*
<https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942>
accessed 30 July 2025

Copyright Act of 1976, 17 USC §§ 101–810

Executive Office of the President, Executive Order No 14110 ‘*Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*’, 88 FR 75191 (2023)
<<https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>> accessed 09 July 2025

Executive Office of the President, Executive Order No 14179 ‘*Removing Barriers to American Leadership in Artificial Intelligence*’, 90 FR 8741 (2025)
<<https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>> accessed 7 July 2025

H.R.5515 - 115th Congress (2017-2018): John S. McCain National Defense Authorization Act for Fiscal Year 2019, H.R.5515, 115th Cong. (2018)
<<https://www.congress.gov/bill/115th-congress/house-bill/5515>> accessed 8 July 2025

H.R.6216 - 116th Congress (2019-2020): National Artificial Intelligence Initiative Act of 2020, H.R.6216, 116th Cong. (2020), <<https://www.congress.gov/bill/116th-congress/house-bill/6216/text>> accessed 8 July 2025

2) Table of Cases

EU Cases

Cases of the CJEU

C-683/17 Cofemel – Sociedade de Vestuário SA v G-Star Raw CV [2019] ECLI:EU:C:2019:721

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

C-604/10 Football Dataco Ltd v Yahoo! UK Ltd [2012] ECLI:EU:C:2012:115

C-469/17 Funke Medien NRW GmbH v Bundesrepublik Deutschland [2019] ECLI:EU:C:2019:623

C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECLI:EU:C:2009:465

Case C-310/17 Levola Hengelo BV v Smilde Foods BV [2018] ECLI:EU:C:2018:899

C-250/25 *Like Company v Google Ireland* (Request for a preliminary ruling, pending, 3 April 2025) <<https://curia.europa.eu/juris/showPdf.jsf?docid=300681&doclang=EN>> accessed 07 August 2025;

C-145/10 *Painer v Standard Verlags GmbH and Others* [2011] ECLI:EU:C:2013:138

Opinion of AG Szpunar, joined Cases C-580/23 and C-795/23 *Mio AB and Others v Galleri Mikael & Thomas Asplund Aktiebolag; konektra GmbH, LN v USM U. Schärer Söhne AG* [2025] ECLI:EU:C:2025:330

Request for a preliminary ruling from the Budapest Környéki Törvényszék (Hungary) lodged on 3 April 2025 – *Like Company v Google Ireland Limited* (Case C-250/25, *Like Company*) (C/2025/3039) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62025CN0250>>, accessed 07 August 2025

C-833/18 *SI and Brompton Bicycle Ltd v Chedech/Get2Get* [2020] ECLI:EU:C:2020:461

EU Member State court cases

No. 310 O 227/23, *Robert Kneschke v. LAION e.V.*, Regional Court of Hamburg, Judgment of 27 September 2024

Rozsudek Městského soudu v Praze sp zn 10 C 13/2023–16 (Czech Republic, 11 October 2023) (Municipal Court, Prague – Judgment No. 10 C 13/2023 16 (11 Oct 2023) https://msp.gov.cz/documents/14569/1865919/10C_13_2023_10/108cad3e-d9e8-454f-bfac-d58e1253c83a accessed 08 August 2025.

US Cases

Supreme Court of the United States

Andy Warhol Foundation for the Visual Arts Inc v Goldsmith, 598 U.S. 508 (2023)

Harper & Row Publishers Inc v Nation Enterprises, 471 US 539 (1985)

Campbell v Acuff-Rose Music Inc, 510 US 569 (1994)

United States Courts of Appeals (Circuit Courts)

Naruto v. Slater 888 F.3d 418 (9th Cir. 2018)

Urantia Foundation v. Maaherra 114 F.3d 955 (9th Cir. 1997)

Thaler v. Perlmutter, No. 23-5233 (D.C. Cir. 2025) <[23-5233.pdf](#)> accessed 2 August 2025

United States District Courts

Allen v Perlmutter, No 1:24-cv-02665 (D. Colo, 2024)

Bartz v. Anthropic PBC, 3:24-cv-05417, (N.D. Cal. 2025)

Disney Enterprises Inc v Midjourney Inc., 2:25-cv-05275, (C.D. Cal.) (Complaint), 2.

Kadrey v Meta Platforms Inc, No 23-cv-03417-VC, Order Denying Plaintiffs' Motion for Partial Summary Judgment and Granting Meta's Cross-Motion for Partial Summary Judgment, (N.D. Cal. 2025) <<https://law.justia.com/cases/federal/district-courts/california/candce/3:2023cv03417/415175/598/>> accessed 05 August 2025

The New York Times Company v Microsoft Corporation and OpenAI Inc, No 1:23-cv-11195, (SDNY, December 2023)

Thaler v Perlmutter, No 1:22-cv-01564-BAH (D.D.C 2023)

Thomson Reuters Enterprise Centre GmbH et al v. ROSS Intelligence Inc., No. 1:2020cv00613, 547 (D. Del. 2023)

3) Other Official Documents

EU Official Documents

European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe, COM/ (2018) 237 final

European Commission, 'Code of Practice for General-Purpose AI Models' (2025) <<https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>> accessed 3 August 2025

European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Coordinated plan on Artificial Intelligence COM (2018) 795 final

European Commission, 'Guidelines on the scope of the obligations for general-purpose AI Models established by AI Act' (2025) <<https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>> accessed 3 August 2025

European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Making the most of the EU's innovative potential: An intellectual property action plan to support the EU's recovery and resilience* COM (2020) 760 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

European Commission, 'Template for the Public Summary of Training Content for General-Purpose AI Models' (2025) <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>> accessed 3 August 2025

European Commission, 'White Paper on Artificial Intelligence – A European approach to excellence and trust', COM (2020) 65, final

European Economic and Social Committee, ‘Opinion on Artificial Intelligence — The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society’ [2017] OJ C 288/01

European Parliament, Resolution of 19 May 2021 on artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI)
<https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_EN.html> accessed 14 June 2025

European Parliament, Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI)
https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html accessed 6 August 2025

European Parliamentary Research Service, ‘AI and copyright: The training of general-purpose AI’, PE 769.585 (2025)
<[https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2025\)769585](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769585)> accessed 1 August 2025.

United States Official Documents

Executive Office of the President, National Science and Technology Council, Committee on Technology ‘Preparing for the Future of Artificial Intelligence’ (2016)
<https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> accessed 7 July 2025

National Science and Technology Council, Select Committee on Artificial Intelligence ‘National Artificial Intelligence Research and Development Strategic Plan’ (2023 Update)
<<https://www.nitrd.gov/pubs/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>> accessed 7 July 2025

U.S. Copyright Office, ‘Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence’, 88 FR 16190, (2023)

4) Research, Studies and Surveys

Bulayenko O and others, ‘AI Music Outputs: Challenges to the Copyright Legal Framework’, reCreating Europe Report (2022) <<https://zenodo.org/records/6405796#.YkbhSihBw2w>> accessed 15 July 2025.

Cooper A. Feder and others, ‘Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models’ (2024) <<https://arxiv.org/abs/2505.12546>> accessed 31 July 2025

Council of the European Union, ‘Revised summary of Member States contributions on the policy Questionnaire on the relationship between generative artificial intelligence and copyright and related rights Prepared by the Hungarian Presidency’ 16710/1/24 REV 1 (2024)
<<https://data.consilium.europa.eu/doc/document/ST-16710-2024-REV-1/en/pdf>> accessed 7 August 2025

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

Copyright data management and artificial intelligence’, Publications Office of the European Union, 2022, 190 <<https://data.europa.eu/doi/10.2759/570559>> accessed 03 August 2025

European Commission, ‘Trends and Developments in Artificial Intelligence — Challenges to the Intellectual Property Rights Framework’ (2020) <<https://digital-strategy.ec.europa.eu/en/library/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-framework>> accessed 14 June 2025

Lucchi N, ‘Generative AI and Copyright: Training, Creation, Regulation’ (Study for the European Parliament, Policy Department for Citizens’ Rights and Constitutional Affairs, 2025) <[https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU\(2025\)774095_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf)> accessed 12 July 2025

Congressional Research Service, Generative Artificial Intelligence and Copyright Law, LSB10922 (2025) <<https://www.congress.gov/crs-product/LSB10922>> accessed 31 July 2025

U.S. Copyright Office, ‘*Copyright and Artificial Intelligence: Part 1: Digital Replicas*’ (2024) <[Copyright and Artificial Intelligence, Part 1 Digital Replicas Report](#)> accessed 11 July 2025

U.S. Copyright Office, ‘Copyright and Artificial Intelligence. Part 2: Copyrightability’ (2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>> accessed 10 August 2025

U.S. Copyright Office, ‘*Copyright and Artificial Intelligence: Part 3 – Generative AI Training*’ (pre-publication version, May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>> accessed 1 August 2025

U.S. Copyright Office, Notice of inquiry and request for comments ‘Artificial Intelligence and Copyright’, 88 FR 59942 (2023) <<https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf>> accessed 01 August 2025

U.S. Copyright Office, Registration Decision Regarding “Théâtre D’opéra Spatial” by Jason M. Allen (Copyright Review Board, 5 September 2023) <<https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf>> accessed 07 August 2025

Bibliography

Articles

Baraniuk C, ‘Computer paints “new Rembrandt” after old works analysis’ (2016) BBC News <<https://www.bbc.com/news/technology-35977315>> accessed 5 May 2025

Dornis T, ‘The Training of Generative AI is Not Text and Data Mining’ (2024) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4993782> 9 accessed 30 July 2025

Goldstein P, Stuetzle C, Bischoff S ‘Kneschke vs. LAION - Landmark Ruling on TDM exceptions for AI Training Data - Part 2’ (2024) Kluwer Copyright Blog <<https://legalblogs.wolterskluwer.com/copyright-blog/kneschke-vs-laion-landmark-ruling-on-tdm-exceptions-for-ai-training-data-part-2/>> accessed 30 July 2025

Krishnan G, ‘Could the Future of Music Composition Be AI? Analysing the Implications of AIVA — An AI Composer’m’ (2022) Medium <https://medium.com/@gaurav_krishnan/could-the-

[future-of-music-composition-be-ai-analysing-the-implications-of-aiva-an-ai-composer-8e3762354860](#)> accessed 5 May 2025

Lee E, 'Fair Use and the Origin of AI Training' (2025) Houston Law Review, Santa Clara Univ. Legal Studies Research Paper No. 5253011 <<https://ssrn.com/abstract=5253011>> accessed 22 August 2025

Lemley M and Casey B, 'Fair Learning' (2021) 99 Texas Law Review, 743

Lemley M, 'How Generative AI Turns Copyright Upside Down' (2024) Science and Technology Law Review <https://law.stanford.edu/wp-content/uploads/2024/09/2024-09-30_How-Generative-AI-Turns-Copyright-Upside-Down.pdf> accessed 3 June 2025

Novelli C and others, 'Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity' (2024) 55 *Computer Law & Security Review* 106066

Samuelson P, 'Allocating Ownership Rights in Computer-Generated Works' (1985) 47 University of Pittsburgh Law Review 1185

Samuelson P, 'Does Training AI on Copyrighted Works Violate Copyright Law?' (2024) 103 Boston University Law Review 1389

Senftleben M, 'The TDM Opt-Out in the EU – Five Problems, One Solution' (2025) Kluwer Copyright Blog <<https://legalblogs.wolterskluwer.com/copyright-blog/the-tdm-opt-out-in-the-eu-five-problems-one-solution/>> accessed 30 July 2025

Villasenor J, 'The problems with California's pending AI copyright legislation' (2025) the Brookings Institution <<https://www.brookings.edu/articles/the-problems-with-californias-pending-ai-copyright-legislation/>> accessed 30 July 2025

White & Case LLP, 'AI Watch: Global Regulatory Tracker – United States' (2025) White & Case <<https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states>> accessed 29 July 2025

Other Online Sources

AI Summit 2025, 'President Trump on Copyright and AI Regulation' (YouTube, 23 July 2025) <<https://www.youtube.com/watch?v=BrTT7dX0mcQ>> accessed 05 August 2025, 20:09–25:46
SACEM, SACEM and Artificial Intelligence (May 2025) <<https://www.createurseditors.sacem.fr/en/news/our-society/sacem-and-artificial-intelligence>> accessed 6 August 2025

Authors Guild, FAQs on the Authors Guild's Positions and Advocacy Around Generative AI (2025) <<https://authorsguild.org/advocacy/artificial-intelligence/faq/>> accessed 6 August 2025

Coalition of European creator organisations, 'Joint statement regarding the AI Act implementation measures adopted by the European Commission' (Brussels, 30 July 2025) <<https://composeralliance.org/news/2025/7/joint-statement-regarding-the-ai-act-implementation-measures-adopted-by-the-european-commission/>> accessed 6 August 2025

DataDome, 'How to Use Your Robots.txt to (Even Partially) Block Bots From Crawling Your Site' (2025) <<https://datadome.co/bot-management-protection/blocking-with-robots-txt/>> accessed 27 July 2025

DeviantArt, 'UPDATE: All Deviations Are Opted Out of AI Datasets' (2022) <<https://www.deviantart.com/team/journal/UPDATE-All-Deviations-Are-Opted-Out-of-AI-Datasets-934500371>> accessed 27 July 2025

European Commission, AI High Level Expert Group, <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>> accessed 15 July 2025

European Writers' Council, 'Joint letter to the European Parliament's JURI Committee on copyright and generative artificial intelligence – opportunities and challenges' (19 June 2025) <https://europeanwriterscouncil.eu/2506_jl_juritoep/?utm_source> accessed 8 August 2025

Federal Trade Commission, 'FTC Announces Crackdown on Deceptive AI Claims and Schemes' (2024) <[FTC Announces Crackdown on Deceptive AI Claims and Schemes | Federal Trade Commission](#)>, accessed 08 July 2025

The Guardian, 'Trump Fires Copyright Office Chief Shira Perlmutter' (12 May 2025) <https://www.theguardian.com/us-news/2025/may/12/trump-fires-copyright-office-shira-perlmutter>> accessed 05 August 2025

Microsoft, 'Prompt Engineering Techniques' Microsoft Learn – Azure AI Foundry (2025) <<https://learn.microsoft.com/en-us/azure/ai-foundry/openai/concepts/prompt-engineering?tabs=chat>> accessed 10 July 2025

New York Post, 'Trump Fires Head of US Copyright Office, Stoking Fears about Deregulation of AI' (12 May 2025) <<https://nypost.com/2025/05/12/business/trump-fires-head-of-us-copyright-office-stoking-fears-about-deregulation-of-ai/>> accessed 05 August 2025

OpenAI, 'Bots on OpenAI Platform' <<https://platform.openai.com/docs/bots>> accessed 27 July 2025

OpenAI, 'Complying with the EU AI Act's Prohibited Practices Requirements' <<https://help.openai.com/en/articles/10658393-complying-with-the-eu-ai-acts-prohibited-practices-requirements>> accessed 3 August 2025

OpenAI, 'Prompting: Overview – OpenAI API' OpenAI Platform Documentation <<https://platform.openai.com/docs/guides/prompting>> accessed 2 July 2025

Stephan Brüggemann, 'Meta Tag Robot: Understanding Its Role in SEO Optimization' (2025) <<https://www.seocopilot.com/meta-tags/metatag-robot-understanding-its-role-in-seo-optimization>> accessed 27 July 2025