

Copyright, the Al Act and Extraterritoriality

João Pedro Quintais, Associate Professor University of Amsterdam, Institute for Information Law (IViR)

Technology has always been a fundamental enabler of creativity. By the same token, any use of technology to automate elements of the creative process has always been met with resistance. This pattern goes back hundreds of years: the internet, digital recording, photography, phonographic recording, chemical synthesis of blue pigments and the printing press.¹

Generative artificial intelligence (AI) brings another, arguably more extreme item to the list.² On the one hand, we are marvelling at its capacity to create new, human-like content. On the other, we are panicking, as we see factory-like automation being extended to the very heart of human expression. In just a few clicks, anyone can generate text, videos, music or images thanks to AI models that are trained on content originally produced by humans. In some ways, "this is a completely new problem that we've been arguing about for 500 years."³ At the same time, there is a genuine novelty in the sheer scale of the issue, as creative works can be produced in a matter of seconds by anyone. This novelty raises challenging questions. Do AI-

generated outputs constitute protected works, derivative works or are they infringing third party works used to train models?

The impact of generative AI on creativity defies assumptions and easy answers. In a recent survey of 10,000 designers, 52% use generative

'The impact of generative AI on creativity defies assumptions and easy answers'

Al tools in their work and more respondents believe Al will increase their revenues (47%) rather than reduce them (33%).⁴ Yet, this optimism is not universal. Industry forecasts from the International Confederation of Societies of Authors and Composers (CISAC) predict a 24% and 21% decline in revenues for music and audiovisual creators, respectively, by 2030.⁵

¹ This policy brief was prepared by the author for the Lisbon Council. The Lisbon Council has received support for this research project from Meta and Google. Any errors of fact or judgement are the Lisbon Council team's and author's sole responsibility. This research was undertaken by the author in academic independence and in accordance with principles, standards and duties set forth in the Netherlands Code of Conduct for Research Integrity (2018).

² But see Arvind Narayanan and Sayash Kapoor, "AI as Normal Technology," Knight First Amendment Institute, 15 April 2025.

³ Benedict Evans, "Generative AI and Intellectual Property," Benedict Evans, 27 August 2023.

^{4 99}designs Team, "Freelance Design in the Age of AI," 99designs, 22 October 2024.

⁵ CISAC and PMP Strategy, Study on the Economic Impact of Generative AI in the Music and Audiovisual Industries (CISAC, November 2024).

The recently adopted AI act⁶ has partly attempted to address some of the key copyright concerns surrounding AI but has left much room for interpretation and legal uncertainty. While explicitly stating that the development of AI models must comply with copyright law, namely the copyright in the digital single market (CDSM) directive,⁷ the AI act has also opened up uncertainties about the interaction between the two legal instruments.

One of the most prominent issues is the extraterritorial reach of the Al act's copyright provisions. Copyright law is inherently territorial and primarily regulates activities – such as Al model training – according to the laws of the place where they occur. In contrast, the Al act regulates any model made available to users in the European Union, while also attempting to extend its reach to training that takes place outside the European Union. This raises a

'One of the most prominent issues is the extraterritorial reach of the AI act's copyright provisions' pressing question: if a model is trained in country outside the European Union in full compliance with that country's copyright legislation but where that legislation differs from European legislation, can it be released in the European Union?

This is not just a technical discussion for lawyers. The resulting uncertainty poses concrete challenges for the development and adoption of Al in Europe.

Furthermore, restrictive copyright rules alone are unlikely to address the broader concerns around fair remuneration for creators. It is therefore crucial to understand the implications of the current legislative provisions and to chart a practical and balanced way forward.

This paper aims to address this challenge by exploring the relationship between the AI act and European copyright law and providing an in-depth analysis of the issue of extraterritoriality. The structure of the analysis is as follows:

- The next section offers an overview of copyright issues across the AI lifecycle, including challenges related to training data, AI models and generated outputs.
- The paper then examines the AI act's copyright provisions, beginning with an overview of the AI value chain and key definitions before examining the legal distinctions between the AI act and European copyright law. It goes on to explore the text-and-data mining (TDM) exceptions in the CDSM directive, and the copyright obligations imposed by the AI act, with a particular focus on the requirement for general purpose AI (GPAI) model providers to respect copyright opt-outs.
- The analysis then turns to the extraterritorial implications of these obligations. It first examines territoriality principles and applicable law in TDM, followed by an assessment of whether the AI act improperly extends its reach beyond European borders. This section closes with a brief discussion of the AI act's GPAI code of practice.

The paper concludes by summarising its key findings and highlighting the broader implications of the Al act's approach to copyright compliance.

⁶ 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) (Text with EEA relevance) (AI Act).

⁷ 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 (Text with EEA relevance.) (CDSM Directive).

Copyright and Al: An Overview of Key Issues

Copyright concerns arise at various stages of the AI lifecycle or value chain, including the input or training data stage, the AI model itself and the outputs generated by or with the assistance of an AI model or system.⁸

At the input stage, training and developing AI models involves practices such as web scraping, data collection and processing, pre-training, fine-tuning and training proper.⁹ These activities often involve reproductions of protected content and fall under the broad legal concept of TDM, as defined in Article 2(2) of the CDSM directive. As explored below, most of the copyright-relevant provisions of the AI act relate to this input or training stage.

With respect to the model itself, key questions centre on the legal status of models' weights as protected databases, whether or not a model memorises copyrighted content and, if so, to what extent this would constitute an unauthorised reproduction under existing copyright laws.¹⁰

In terms of output, key issues include whether Al-generated outputs qualify for copyright protection, constitute derivative works or infringe upon third-party works used in the training of a model.¹¹ Within the European Union, a notable and evolving issue is the extent to which copyright exceptions apply to Al-generated outputs, including those based on freedom of expression, such as quotation, criticism, review, caricature, parody and pastiche.¹²

Additional complexities are introduced by the intersection of copyright law with private contractual agreements, such as Al provider's terms and conditions, which can affect authorship, ownership and enforcement.¹³

Finally, an increasingly urgent policy concern is how creators should be remunerated for the use of their works in generative AI models.¹⁴

The AI "Value Chain" or "Lifecycle," and Key Terms

The Al act is a complex and detailed piece of legislation.¹⁵ Chapter V specifically addresses GPAI models and outlines a set of copyright-related responsibilities,¹⁶ which are due to take effect in August 2025, with enforcement expected only from August 2026.¹⁷

⁸ Andrés Guadamuz, "A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs," GRUR International, vol. 73, no. 2, 2024, pp. 111–127; João Pedro Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107.

⁹ Often, these stages are jointly referred to as "training." See e.g., European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025).

¹⁰ A. Feder Cooper and others, "Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models," arXiv, 18 May 2025; A. Feder Cooper and James Grimmelmann, "The Files Are in the Computer: On Copyright, Memorization, and Generative AI," arXiv, 18 July 2024; Tim W. Dornis, "The Training of Generative AI Is Not Text and Data Mining," European Intellectual Property Review, no. 2, 2025; Ivo Emanuilov and Thomas Margoni, "Forget Me Not: Memorisation in Generative Sequence Models Trained on Open Source Licensed Code," Zenodo, February 2024; European Union Intellectual Property Office, Development of Generative Artificial Intelligence from a Copyright Perspective (Alicante: European Union Intellectual Property Office); Nuno Sousa e Silva, "Are AI Models' Weights Protected Databases?," Kluwer Copyright Blog, 18 January 2024.

¹¹ The latter issue, often linked to memorisation, should instead be viewed from the perspective of regurgitation, extraction, and reconstruction during output generation, as highlighted by A. Feder Cooper and James Grimmelmann, "The Files Are in the Computer: On Copyright, Memorization, and Generative AI," *arXiv*, 18 July 2024. See also Matthew Sag, "Fairness and Fair Use in Generative AI," *Fordham Law Review*, vol. 92, 2024, p. 1887.

¹² NB that the upcoming Court of Justice of the European Union (CJEU) judgement in *Pelham II* may further define "pastiche" as an autonomous EU legal concept, potentially impacting the treatment of AI outputs.

¹³ See e.g., Oleksandr Bulayenko and others, "Al Music Outputs: Challenges to the Copyright Legal Framework," Social Science Research Network, 28 February 2022; Gabriele Cifrodelli and Lilian Edwards, "Copyright and Generative AI: What Can We Learn from Model Terms and Conditions?," Kluwer Copyright Blog, 24 July 2024; Lilian Edwards and others, "Private Ordering and Generative AI: What Can We Learn From Model Terms and Conditions?" CREATE, 24 May 2024.

¹⁴ Christophe Geiger and Vincenzo Iaia, "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI," *Computer Law & Security Review*, vol. 52, 2024, p. 105925; João Quintais, "Generative AI, Copyright and the AI Act," *Computer Law & Security Review*, vol. 56, 2025, p. 106107; Martin Senftleben, "Generative AI and Author Remuneration," *IIC – International Review of Intellectual Property and Competition Law*, vol. 54, 2023, pp. 1535–1560.

¹⁵ On the larger debate of complex legislation and its perils, see Lisa Burton Crawford, "The Problem of Complex Legislation," Legal Theory, 2024, pp. 1–21.

¹⁶ See also supporting Recitals 104 to 109 AI Act.

¹⁷ Art. 113 Al Act.

One helpful way to understand the Al act is through the concept of the Al "value chain" or "lifecycle." In simple terms, this refers to the end-to-end process of developing, deploying and managing Al systems. It is an ongoing cycle that includes stages such as data collection, data processing, model training, evaluation, deployment, monitoring and governance.

To navigate the AI value chain and its copyright implications, it is important to first clarify key terms as defined by the act. These include AI systems,¹⁸ AI models,¹⁹ GPAI systems²⁰ and GPAI models.²¹ Importantly, the potential breath and scope of these definitions have already

'To navigate the AI value chain and its copyright implications, it is important to first clarify key terms as defined by the act' led to Commission-issued guidelines on the definition of an AI system as well as a targeted consultation that will contribute to the upcoming Commission guidelines on GPAI models, including different facets of their definition.²²

Within the framework of the Al act, GPAl models are distinct from Al systems. GPAl models are highly flexible tools capable of performing multiple tasks, often trained on large datasets

using techniques such as self-supervised or reinforcement learning. They can be accessed via application programming interfaces (APIs) or as direct downloads and may be modified or integrated into larger AI systems. AI systems, on the other hand, incorporate additional elements, such as user interfaces,²³ that allow users to interact with the underlying models.²⁴

The Al act introduces specific rules for GPAI models, particularly those that pose systemic risks. These rules apply when models are released on the market and when they are integrated into Al systems. However, models used solely for internal processes or research before their commercial release are largely exempt from these obligations.²⁵

For instance, generative pre-trained transformers (GPTs)²⁶ are examples of GPAI models, whereas applications like ChatGPT, Midjourney, DALL-E and Firefly are AI systems that incorporate such models. While the AI act does not explicitly define "generative AI," it does clarify that large generative models – those that produce text, audio, images or video – fall into the category of GPAI models.²⁷

¹⁸ Art. 3(1) AI Act.

¹⁹ This term is not explicitly defined in Article 3 AI Act.

²⁰ Art. 3(66) AI Act. See also European Commission, Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act), C(2025) 924 final (6 February 2025).

²¹ Art. 3(63) AI Act.

²² European Commission, Guidelines on the Definition of an Artificial Intelligence System Established by the AI Act (6 February 2025); European Commission, Targeted Consultation in Preparation of the Commission Guidelines to Clarify the Scope of the Obligations of Providers of General-Purpose AI Models in the AI Act (22 April 2025).

²³ Recital 97 AI Act.

²⁴ For additional details on the distinction, see João Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107.

²⁵ João Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107; João Pedro Quintais, "What Is a 'Research Organisation' and Why It Matters: From Text and Data Mining to AI Research," GRUR International, vol. 74, no. 5, 2025, pp. 397–398; Michael Veale and João Pedro Quintais, "The Obligations of Providers of General-Purpose AI Models" in The Artificial Intelligence Act – A Thematic Commentary, ed. by Gabriela Zanfir Fortuna and others (Hart, 2025).

²⁶ Examples include OpenAI's GPT-4, Google DeepMind's Gemini 1.5 and Meta's Llama 2.

²⁷ Recitals 99 and 105 AI Act. Note that the AI Act also regulates non-generative GPAI models, such as those used in recommender systems.



These distinctions are particularly important because the Al act's copyright-related obligations apply to GPAI model providers rather than those who build or deploy Al systems. Organisations involved earlier in the value chain, such as those compiling datasets or scraping online content, are not classified as GPAI model providers – even if they engage in TDM

of copyrighted material. As a result, they are not subject to the Al act's copyright obligations.

One of the main challenges of Al systems is that they are not static, but rather are constantly evolving. As a result, legal and policy issues, in particular copyright 'One of the main challenges of AI systems is that they are not static, but rather are constantly evolving'

concerns, can arise at any point in their development. As noted, these complexities range from the data used for training and the AI model itself to the content it produces. As AI technologies advances, the frameworks used to analyse them must also adapt, requiring experts to examine each stage with ever greater precision.



The Intricate Interplay of the AI Act and EU Copyright Law

One of the pivotal provisions of the Al act is Article 53, which establishes, inter alia, copyrightrelated obligations for providers of GPAI models. These obligations require compliance with EU copyright law, including the opt-out requirement outlined in Article 4(3) of the CDSM directive.

However, a fundamental challenge arises: the Al act and European copyright law are rooted in distinct legal frameworks and operate under different enforcement and remedial mechanisms.²⁸ Given these inherent differences, how can their relationship be coherently defined when they operate in separate domains?²⁹

The Al act is fundamentally a public law instrument, structured through a product safety lens to serve the public interest by imposing systemic compliance obligations on certain providers. In contrast, copyright law operates within a private law framework, conferring exclusive or remuneration rights to rights holders and enabling individual enforcement.

Adding further complexity to this relationship, Recital 108 of the Al act asserts that the regulation does not interfere with the enforcement of copyright rules, while Recital 109 clarifies that compliance with the Al act does not override European copyright law. However, Article 2 of the Al act, which defines the act's scope and exclusions, makes no explicit reference to copyright.³⁰ This omission introduces legal uncertainty, particularly for Al developers and providers seeking clarity as to whether and how they must address their copyright-related obligations in the Al act.

The complexity deepens when considering the interaction between the AI act and the TDM exceptions outlined in the CDSM directive. As the directive allows varying degrees of implementation across 27 European member states, national discrepancies in copyright law exist, creating an uneven legal landscape.

TDM in the European Union is regulated by Articles 2(2), 3 and 4 of the CDSM directive, which define the concept and establish two distinct copyright exceptions:

- Article 2(2) defines TDM broadly as "any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations."
- Article 3 applies to TDM conducted for scientific research by "research organisations" and "cultural heritage institutions," provided they have lawful access to the works or subject matter and meet certain additional conditions.³¹
- Article 4 provides an exception for reproductions and extractions of lawfully accessed works or subject matter for the purposes of TDM. This exception is subject to rights reservation by rights holders, including through "machine-readable means in the case

²⁸ Alexander Peukert, "Copyright in the Artificial Intelligence Act - A Primer," GRUR International, vol. 73, no. 6, 2024, pp. 497-509.

²⁹ Séverine Dusollier and others, "Copyright and Generative AI: Opinion," JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law, vol. 16, no. 1, 2025.

³⁰ By contrast, liability issues for intermediaries, addressed in Chapter II of the digital services act, are explicitly excluded in Art. 2(5) AI Act.

³¹ Thomas Margoni and Martin Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," GRUR International, vol. 71, no. 8, 2022, pp. 685–701. On the requirement of lawful access, see e.g., Thomas Margoni, "TDM and Generative AI: Lawful Access and Opt-Outs," Auteurs&Media, 2024; Kacper Szkalej, The Paradox of Lawful Text and Data Mining? Some Experiences from the Research Sector and Where We (Should) Go from Here, GRUR International, Volume 74, Issue 4, April 2025, Pages 307–319, <u>https://doi.org/10.1093/grurint/ikaf029</u>.

of content made publicly available online," for instance using metadata and terms and conditions of a website or a service. This possibility of rights reservation is commonly referred to as the "opt-out" requirement.³²

A fundamental question arises from the broad definition of TDM in Article 2(2), and the exceptions in Articles 3 and 4: how does the TDM definition apply to the input and training stages of developing GPAI models?

This appears to be clarified by Recital 105 of the Al act, which notes that generative Al models offer significant opportunities for innovation while simultaneously posing challenges to artists, authors and other creators in relation to the creation, distribution, use and consumption of their work. It acknowledges that TDM techniques for these models often require extensive use of copyright-protected materials, which requires authorisation from rights holders, unless exceptions under Articles 3 and 4 of the CDSM directive apply. Moreover, Recital 105 underscores the obligation of GPAI model providers to adhere to the rights reservation mechanism outlined in Article 4(3) where relevant.

This suggests that the AI act views the development of GPAI models as inherently involving acts of TDM. When this process involves copyright-protected content – which is usually the case – it must comply with the CDSM directive's TDM exceptions or pre-existing exceptions for temporary reproductions or research in the copyright *acquis*.³³

However, legal scholars disagree on this point. Some argue that all copyright-relevant reproductions and extractions during Al training qualify as TDM.³⁴ Others exclude certain acts of reproduction from TDM exceptions,³⁵ while some suggest that generative Al model training falls outside the TDM definition altogether.³⁶

One thing is clear: TDM exceptions do not give AI developers the right to publicly share or distribute the results of their data mining if they contain copyrighted material. The exceptions apply only to the processes of copying and extracting protected content, not to making that content publicly available.

For example, the exceptions do not permit datasets created through TDM to be made public *if the datasets contain copies of protected content*.³⁷ Similarly, any copies made during the TDM process that do not meet the conditions set out in Articles 3 and 4 of the CDSM directive are not covered by the exceptions.

³² NB such reservation shall not affect the application of the TDM exception for scientific purposes in Art. 3 CDSM Directive. On this provision, see e.g., Giuseppe B. Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 479–487; Sévérine Dusollier and others, "Copyright and Generative AI," *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 16, no. 1, 2025; Hanjo Hamann, "Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and Their (In)Compatibility with Article 4(3) of the Copyright DSM Directive," *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 15, no. 2, 2024; Péter Mezei, "A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 461-469.

³³ NB Art. 3 CDSM Directive may be combined with the optional exception covering uses for non-commercial scientific research purposes in Art. 5(3)(a) Directive 2001/29/EC (InfoSoc Directive), which already covered certain TDM activities. Likewise, certain temporary acts of reproduction covered by Arts. 3 and 4 CDSM Directive would already be exempted under the mandatory exception for transient and temporary copying in Art. 5(1) InfoSoc Directive. These overlapping exceptions are explicitly recognised in recitals 9 and 15 CDSM Directive.

³⁴ Giuseppe B. Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," European Intellectual Property Review, vol. 46, no. 7, 2024, pp. 479–487.

³⁵ Eleonora Rosati, "Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law," European Journal of Risk Regulation, 2024.

³⁶ Tim W. Dornis, "The Training of Generative AI Is Not Text and Data Mining," European Intellectual Property Review, no. 2, 2025.

³⁷ See Art. 3(2) ("Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results") and Art. 4(2) ("Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining") CDSM Directive.



Copyright in the Al Act

Building on this background, the AI act sets out two key copyright-related obligations for all GPAI model providers:

- Compliance policy: Article 53(1)(c) mandates that GPAI model providers implement a policy to ensure compliance with European copyright law. In particular, providers must identify and respect, including through the use of state-of-the-art technologies, reservations of rights (i.e., "opt-outs") expressed under Article 4(3) of the CDSM directive.
- Training data disclosure: Article 53(1)(d) requires these providers to draft and publicly disclose a sufficiently detailed summary of the content used to train their GPAI models. This summary must follow a template provided by the European AI Office.

These two obligations raise common uncertainties, particularly regarding who else in the AI development process must comply with them. While they explicitly apply to GPAI model providers, it remains unclear whether other actors in the AI lifecycle are also affected.

A crucial distinction can be drawn between "upstream providers," who handle web scraping, crawling and supplying training tools and datasets, and "downstream providers," which include AI system providers.³⁸

Recital 97 clarifies that when an AI model is directly integrated into an AI system developed by the same provider, both the model and the system must comply with the AI act. However, when models and systems come from different entities, this obligation does not seem to apply,³⁹ despite the documentation obligations under Article 53(1)(b) remaining relevant.⁴⁰

This raises a key question: when separate entities, rather than a single company, are responsible for model and system development and deployment, how should responsibility for ensuring compliance with copyright regulations be assigned? Here, the situation becomes much less clear.

The implications of the distinction between upstream, model and downstream providers are as follows:

- Upstream providers that engage in TDM but do not qualify as GPAI model providers remain subject to Articles 3 and 4 of the CDSM directive but are generally excluded from the obligations set forth in the AI act.
- Downstream AI system providers that integrate models are, in principle, not subject to the obligations imposed on model providers under Article 53 unless they are vertically integrated.

³⁸ Note that some of these downstream providers could potentially also be scraping or crawling additional information as part of processing queries. One example is retrieval-augmented generation (RAG), a technique that enhances AI-generated responses by retrieving relevant external information from a database or the web before generating text, improving accuracy and relevance. This paper does not further examine the copyright implications of RAG techniques. On the topic of RAG and copyright, see e.g., European Union Intellectual Property Office (EUIPO), *Development of Generative Artificial Intelligence from a Copyright Perspective* (EUIPO, 2025), noting that with RAG, copyright-protected content is not only used for training, but also for content generation purposes. See also United States Copyright Office, Copyright and Artificial Intelligence, Part 3: Generative AI Training (Pre-Publication Version), pp. 1–113 (USCO, 2025).

³⁹ Alexander Peukert, "Copyright in the Artificial Intelligence Act – A Primer," GRUR International, vol. 73, no. 6, 2024, pp. 497–509.

⁴⁰ See also Recital 118 AI act, which seeks – albeit somewhat unclearly – to regulate the interplay between the AI act and the digital service act. On this topic, see, e.g., Giuseppe B. Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 479–487; Paddy Leerssen, "Embedded GenAI on Social Media: Platform Law Meets AI Law – DSA Observatory', DSA Observatory Blog, 16 October 2024; João Pedro Quintais, "Generative AI, Copyright and the AI Act," *Computer Law & Security Review*, vol. 56, 2025, p. 106107.



Furthermore, the AI act and the CDSM directive operate independently. This means that if a GPAI model provider does not meet its obligations under the AI act, this non-compliance does not automatically amount to copyright infringement, although it could still lead to administrative fines under the AI act.⁴¹

However, the situation becomes even more complex in cases where non-compliance with the AI act could have copyright implications. For example, if a GPAI model provider fails to employ measures (like web crawlers) that follow instructions expressed in a standard technical format for opt-outs specified in the GPAI code of practice (and approved by the AI Office), a national court might interpret this as a failure to respect the rights reservation mechanism under Article 4(3) of the CDSM directive. Should that be the case, the provider's TDM activities may no longer fall under the copyright exceptions, potentially leading to infringement claims from rights holders.

While both copyright obligations in the Al act carry significant consequences, this paper will primarily focus on the first – the European copyright compliance policy obligation – and its implications for extraterritoriality.

Policies to Respect Copyright, Especially the Opt-out

The exact scope of GPAI model providers' obligation to establish a copyright compliance policy under Article 53(1)(c) of the AI act remains unclear. Must providers ensure full compliance, make reasonable efforts or simply have a formal policy in place without strict enforcement? While interpretations vary, the wording of the AI act suggests that some level of compliance is expected.⁴²

⁴¹ On supervision and enforcement of obligations imposed under the AI Act, see e.g., Arts. 88, 93 and 101. In short, Art. 88 gives the European Commission authority to enforce copyright obligations for GPAI model providers, with tasks delegated to the AI Office. These public bodies can implement co-regulatory measures like codes of practice and standards, and take actions such as restricting or withdrawing non-compliant models (Art. 93). For negligent or intentional violations, Art. 101 allows the Commission to fine providers up to 3% of global annual turnover or €15 million, whichever is higher. Fines take into account any commitments made under Art. 56. Private enforcement is not allowed; individuals may only file complaints with the AI Office.

⁴² See in particular Recital 106 emphasising compliance with the reservation of rights under Art. 4(3) CDSM Directive.



This requirement applies broadly to European copyright law and is not necessarily limited to AI model training. In theory, it could extend across the entire AI value chain, including the moderation of AI-generated outputs.⁴³

For instance, if an Al model generates text, images or music based on copyrighted materials in its training data, its outputs could resemble or incorporate elements of those works. To mitigate potential copyright issues, Al providers might be required to filter and screen outputs – similar to how social media platforms automatically detect and block copyrighted music or videos. In practice, some providers are already deploying these types of measures in their services.⁴⁴

This approach, which has potentially chilling effects to freedom of expression, is already reflected in the GPAI code of practice's measures aimed at mitigating the risk of production of copyright-infringing output by downstream AI systems that integrate generative GPAI models.⁴⁵ It remains unclear, however, whether such downstream filtering is actually required under Article 53(1)(c), and how the deployment of such output moderation tools will interact with the AI act's labelling requirements for synthetic content.⁴⁶

This uncertainty leaves open questions about the extent to which copyright compliance obligations should go beyond the training phase.

Another challenge emerges: if most copyright-infringing outputs occur only after a GPAI model has been integrated into an Al system, how can compliance be ensured beyond the training phase? This issue becomes even more complex when the model and system providers are separate entities.

The core aim of the Al act's policies obligation on GPAI model providers is to ensure that (appropriate) opt-outs under Article 4(3) of the CDSM directive are identified and respected using state-of-the-art technologies. In practice, this means that these providers must implement technical measures that allow rights holders to prevent their works from being used in Al model training. The mention of "state-of-the-art technologies" – absent in the CDSM directive – suggests that technical standards for opt-out implementation may be required.⁴⁷

Under Article 4(3) of the CDSM directive, TDM is permitted by default unless rights holders actively opt out "in an appropriate manner, such as machine-readable means." Recital 18 clarifies that, for publicly available online content, the appropriate way to reserve rights is through machine-readable methods, including metadata and website terms and conditions. In other cases, rights reservations can be made through contractual agreements or unilateral declarations. Additionally, rights holders should be able to implement measures ensuring their opt-outs are respected.

⁴³ Alexander Peukert, "Copyright in the Artificial Intelligence Act – A Primer," GRUR International, vol. 73, no. 6, 2024, pp. 497–509; João Pedro Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106-107.

⁴⁴ European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025). The EUIPO study notes that generative AI providers are implementing tools and techniques – such as content comparison, output filters, prompt rewriting, model editing, and unlearning – to prevent copyright infringement and address issues post-deployment. Many also offer legal indemnification to protect customers.

⁴⁵ GPAI Code of Practice (3rd draft), Measure 1.2.5. (Mitigate the risk of production of copyright-infringing output), para (1)(a): "(1) In order to mitigate the risk that a downstream AI system, into which a generative [GPAI] model is integrated, repeatedly generates output that infringes copyrights or related rights as protected according to Union law on copyright and related rights, Signatories will... (a) make reasonable efforts to mitigate the risk that a model memorizes copyrighted training content to the extent that it repeatedly produces copyright-infringing outputs." On the related freedom of expression concerns, see Caroline De Cock, "From Upload to Output Filters? How the AI Act's Code of Practice Could Threaten Freedom of Expression and Thought," *Tech Policy Press*, 2025.

⁴⁶ On which, see Art. 50(2) Al Act.

⁴⁷ This is confirmed by Art. 53(4) AI Act. See also GPAI Code of Practice (3rd draft), Measure I.2.3.

A recent study by the European Union Intellectual Property Office (EUIPO) distinguishes between two types of opt-out measures, namely:

- "legally-driven" opt-out measures (e.g., unilateral declarations by rights holders, licensing constraints and website terms and conditions) and
- "technically-driven" opt-out measures, including the robots exclusion protocol, TDM reservation protocols, tools adapted for content provenance and authenticity and broader copyright management solutions still in development.⁴⁸

In my view, it is difficult to see how, at this moment, legally-driven measures would meet the requirements of Article 4(3) for TDM of "publicly available online content." Technicallydriven measures, for their part, are more clearly aligned with the requirements of Article 4(3) for TDM of such content. Such measures are therefore the most relevant for our purposes, although it is doubtful whether all of those listed above meet the legal requirements.

Technically-driven measures can be categorised as either location-based – linked to the specific place where the content is hosted or found online – or asset-based, meaning they

are tied directly to the content itself, regardless of its online location.⁴⁹ Examples of technical measures already in use that serve an opt-out function include opt-out tools by Al model providers like Google (Google-Extended),⁵⁰ the TDM reservation protocol,⁵¹ and third-party solutions like Spawning Al's tools (e.g., the Do Not Train Tool Suite).⁵²

'When data sources originate from multiple countries, which copyright law governs the opt-outs?'

Despite these measures, "no single opt-out mechanism has emerged as the sole standard used by rights holders"⁵³ and legal uncertainties remain regarding how the opt-out requirement applies to GPAI model training, including:⁵⁴

- Are location-based, asset-based or hybrid opt-outs all consistent with Article 4(3) of the CDSM directive? How should potential conflicts between different location- and asset-based opt-out mechanisms be dealt with?
- How should conflicts on copyright ownership and overlapping rights in the context of optouts be dealt with? How should the substantive and contractual fragmentation of rights across different territories be dealt with, including differences in assignments and licenses that may affect the lawful accessibility of content for TDM purposes?

⁴⁸ European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025). Examples of the latter include the Liccium Trust Engine Infrastructure or Valuenode's Open Rights Data Exchange platform.

⁴⁹ European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025).

⁵⁰ See these and other examples provided in Paul Keller and Zuzanna Warso, Defining Best Practices for Opting out of ML Training (Open Future Foundation, 2023); João Pedro Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107.

⁵¹ W3C Community Group, TDM Reservation Protocol (TDMRep) (W3C Community Group, 2024).

⁵² Other third-party solutions, such as Cloudflare's crawl blocker – which allows content owners to block AI crawlers and prevent their works from being mined – perform a similar function but may not be considered opt-out tools within the meaning of Art. 4(3) CDSM Directive. See European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025); João Pedro Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107. The EUIPO study considers that these measures fail to meet "the legal criteria for right reservation."

⁵³ European Union Intellectual Property Office (EUIPO), Development of Generative Artificial Intelligence from a Copyright Perspective (EUIPO, 2025).

⁵⁴ Addressing these different questions, see e.g., Guseppe B. Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 479–487; Hanjo Hamann, "Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and Their (In)Compatibility with Article 4(3) of the Copyright DSM Directive," *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 15, no. 2, 2024; Péter Mezei, "A Saviour or a Dead End? Reservation of rights in the age of generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 461-469; João Pedro Quintais, "Generative AI, Copyright and the AI Act," *Computer Law & Security Review*, vol. 56, 2025, p. 106107.

- When can rights holders exercise an opt-out? Can opt-outs be enforced at any stage of Al model development (before, during and after pre-training and training), or do they only apply when content is publicly available online and before it is mined (e.g., during web scraping)?
- For location-based opt-outs, at what level should they be implemented? Should they apply at the webpage level (e.g., via a website's robots.txt file), at a more granular level within the page (e.g., using robots meta tags or elements) or at the level of specific training datasets? When data sources originate from multiple countries, which copyright law governs the optouts?
- What, if any, is the effect of a valid opt-out on the assessment of the lawful access requirement for subsequent TDM of that content?
- Can an opt-out expressed in natural language within a website's terms and conditions constitute an "adequate" machine-readable opt-out under the law?

Most of these questions are far from having simple answers and are likely to keep courts occupied for years to come. For instance, the final question regarding the adequacy of natural language reservations is already being examined in national courts in Germany and The Netherlands, most notably in the judgements of *LAION v. Robert Kneschke* before the Hamburg Regional Court in Germany,⁵⁵ and *DPG Media et al. v. HowardsHome* before the Amsterdam District Court in The Netherlands.⁵⁶

Extraterritoriality and its Discontents

One notable aspect of the obligation to establish "policies to respect copyright" is its potential extraterritorial effect. Recital 106 of the AI act states that this policies obligation applies even if the relevant TDM activities occur outside the European Union.

The goal of this recital is to prevent regulatory arbitrage and ensure a fair competitive environment for GPAI model providers. The idea is that no provider should be able to gain an advantage in the European market by operating under lower copyright standards than those required within the European Union. This reflects a product safety approach embedded in the AI act: if a model is made available in the European Union, it must comply with European law, regardless of where it was developed or trained.

However, this creates tensions with the principle of territoriality, a fundamental concept in both international and European copyright law.⁵⁷ Can – or should – European copyright rules extend beyond its borders in this way?

This section first discusses the principle of territoriality in copyright, jurisdiction and applicable law for TDM. It then critically examines the main issues with the AI act's proposed

⁵⁵ Hamburg District Court, LAION v Robert Kneschke, 310 O.22723, 27 September 24. See e.g. Paul Goldstein, Christianne Stuetzle, and Susan Bischoff, "Kneschke vs. LAION - Landmark Ruling on TDM Exceptions for AI Training Data – Parts 1 and 2," Kluwer Copyright Blog, 13 and 14 November 2024.

⁵⁶ Amsterdam District Court, DPG Media et al. v. HowardsHome, C/13/737170 / HA ZA 23-690, 30 October 2024, ECLI:NL:RBAMS:2024:6563, para 4.33. See generally Etienne Valk and Iris Toepoel, "DPG Media et al vs. HowardsHome – A National Ruling on DSM's Press Publishers' Rights and TDM Exceptions," Kluwer Copyright Blog, 16 January 2025.

⁵⁷ In international copyright law, see e.g., Art. 5(2) Berne Convention and Annette Kur and Ulf Maunsbach, "Choice of Law and Intellectual Property Rights," Oslo Law Review, vol. 6, no. 1, 2019, pp. 43–61. In EU law, see e.g., C-192/04 – Lagardère, ECLI:EU:C:2005:475, 46.

copyright extraterritoriality approach, ending with a brief mention of the GPAI code of practice.

Territoriality, Jurisdiction and Applicable law for TDM

The fundamental principle of territoriality in copyright law means that copyright protection is determined and enforced within the legal framework of a specific country. Each country has its own copyright laws and protection does not automatically extend beyond national borders.⁵⁸ In general, states do not have the authority to regulate conduct beyond their own territory.

Extraterritoriality, on the other hand, refers to the limited circumstances in which a country's copyright laws may have an impact beyond its borders. The principle of territoriality is important in determining both jurisdiction – which country's court has the authority to hear a case – and applicable law – which country's copyright rules apply to the dispute.

In European law, the rules on general and specific jurisdiction are set out in the Brussels Regulation, while the rules on applicable law for non-contractual obligations (e.g. infringement) are laid down in the Rome II Regulation.⁵⁹ Specifically, Article 8 of the Rome II Regulation clarifies that the *lex loci protectionis* – the law of the country for which protection is claimed – applies to copyright infringement, covering both the requirements for protection and its scope.⁶⁰

A clear example of how jurisdiction and applicable law work is the *Hejduk* case. An Austrian photographer sued a German website in an Austrian court for copyright infringement because the content was accessible in Austria. The Austrian court ruled that it had jurisdiction over the case since the alleged infringement affected users in Austria. However, because copyright law is territorial and there is no unitary European copyright title, Austrian copyright law would apply only to damages incurred in Austria, while German copyright law would apply to any damages incurred in Germany.⁶¹

The principle of territoriality and the rule of *lex loci protectionis* then imply that it is crucial to determine where a restricted and potentially infringing act – such as reproducing a work for TDM – takes place to assess which copyright laws apply. If TDM activities, such as copying and extracting data to train Al models, occur outside the European Union, they are subject to the copyright laws of the country where they occur, not European law. Therefore, if a GPAI model is trained in a country outside the European Union, European copyright rules, including TDM opt-out requirements of Article 4 of the CDSM directive, do not apply to the training activities. Once the model is trained and subsequently introduced to the European market, the TDM rules in Article 4 no longer apply, meaning there is no violation of that provision.

⁵⁸ Paul Goldstein and P. Bernt Hugenholtz, International Copyright: Principles, Law, and Practice, 4th ed. (Oxford: Oxford University Press, 2019).

⁵⁹ See, respectively, Regulation (EU) No. 1215/2012 of the European Parliament and of the Council of 12 December 2012 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters (recast), OJ EU 2012 L 351/1, and Regulation (EC) No. 864/2007 of the European Parliament and of the Council of 11 July 2007 on the law applicable to non-contractual obligations (Rome II), OJ 2007 L 199. On the topics of jurisdiction and applicable law for copyright under the Brussels Ibis Regulation and the Rome II Regulation, see Mireille M. M. van Eechoud, "Territoriality and the Quest for a Unitary Copyright Title," *IIC – International Review of Intellectual Property and Competition Law*, vol. 55, no. 1, 2024, pp. 66–88; Annette Kur and Ulf Maunsbach, "Choice of Law and Intellectual Property Rights," *Oslo Law Review*, vol. 6, no. 1, 2019, pp. 43–61; Christian Heinze and Cara Warmuth, "Intellectual Property and the Brussels Ibis Regulation, pp. 147–171 (Edward Elgar Publishing, 2020).

⁶⁰ Malte Stieper and Michael Denga, The International Reach of EU Copyright Through the AI Act (Institut für Wirtschaftsrecht, 2024). The situation is similar in international law, as observed by Annette Kur and Ulf Maunsbach, "Choice of Law and Intellectual Property Rights," Oslo Law Review, vol. 6, no. 1, 2019, pp. 43–61. ("...for all kinds of CJEU, including copyright, applying lex protectionis conforms to a (nearly) universally observed practice, at least in infringement disputes.")

⁶¹ Vienna Commercial Court, Pez Hejduk v EnergieAgentur.NRW GmbH, C-441/13, 22 January 2015, ECLI:EU:C:2015:28.

An example from Professor Peukert illustrates this issue well.⁶² While this is a theoretical case, real-life scenarios may involve significant uncertainty about the location of TDM activities – such as the servers from which content is crawled or scraped – and the example nevertheless brings the point across.

A company trains an AI model in Japan, conducting all TDM activities there. Once trained, the model is made available within the European Union, either on its own or as part of an AI system. A European rights holder wants to sue the provider, arguing that the model was trained without respecting opt-outs under European copyright law for content in the rights holder's repertoire.

This raises two key questions. First, does a national court within the European Union have jurisdiction over the case? Second, which country's copyright law applies?

Under current copyright rules, the answer to the first question is uncertain because the alleged infringing TDM activities took place outside the European Union.⁶³ As for the second question, even if a European court accepts the case, the relevant legal framework for the training process would be Japanese copyright law, not European copyright law. Since Japan's copyright laws on TDM differ from those in the European Union, the likely outcome is that the AI provider did not violate copyright law when training the model, even if under European law, the training of the model might fail to meet the requirements of Article 4 of the CDSM directive.⁶⁴

The same logic applies to any country with less restrictive TDM laws than the European Union.

However, model providers should proceed with caution.

On one hand, European copyright law offers greater legal certainty than many national frameworks. Articles 3 and 4 of the CDSM directive, despite their flaws, set relatively clear conditions for AI development activities involving TDM.⁶⁵ Given the broad interpretation of the reproduction right in both international and national copyright laws, many of these activities likely fall within its scope. In contrast, national laws based on open-ended exceptions may expose model providers to greater legal uncertainty or liability. This is evident in ongoing litigation in the United States, where courts are testing the limits of fair use in AI model development.⁶⁶

On the other hand, as discussed further below, efforts to localise all relevant TDM activities in the development of GPAI models would be far from straightforward. In some cases, a sufficient connection to the European Union may exist, triggering the application of European law. While a motivated party might mitigate this risk, it remains a key concern given the broad scope of the TDM exception and the complexity of Al value chains.

⁶² Alexander Peukert, "Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training," Social Science Research Network, 25 July 2024.

⁶³ Art. 7(2) of the Brussels I Regulation provides special jurisdiction for certain disputes, including torts such as copyright infringement. Jurisdiction lies with the courts of the place where the harmful event occurred or may occur. In case C-441/13 – *Hejduk*, ECLI:EU:C:2015:28, the Court held that a court has jurisdiction based on the place where the damage occurred to hear a copyright infringement claim for online content accessible within its territory, but only for damages incurred within that member state.

⁶⁴ Alexander Peukert, "Copyright in the Artificial Intelligence Act – A Primer," *GRUR International*, vol. 73, no. 6, 2024, pp. 497–509. On Japanese copyright law, see Tatsuhiro Ueno, "Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication," *GRUR International*, vol. 70, no. 2, 2021, pp. 145–52.

⁶⁵ Testing some of the noted legal uncertainties under EU law, see preliminary reference to the CJEU in case C-250/25, Like Company v Google Ireland Limited from the Budapest Környéki Törvényszék (Hungary). For early comment, see Andres Guadamuz, "First Case on AI and Copyright Referred to the CJEU," TechnoLlama, 27 May 2025; Eleonora Rosati, "CJEU Receives First Referral on Chatbots and Copyright," The IPKat, 26 May 2025; Paul Keller, "Do AI Models Dream of Dolphins in Lake Balaton?" Kluwer Copyright Blog, 28 May 2025.

⁶⁶ Chat GPT Is Eating the World (2024) <u>https://chatgptiseatingtheworld.com</u>.



Issues with Extraterritoriality

In defence of extraterritoriality, some authors consider Recital 106 as a "market entry requirement" that derives from the general rule in Article 2(1)(a) of the AI act. According to this provision, the regulation applies, inter alia, to providers placing GPAI models on the market in the European Union, regardless of whether those providers are established or located within the European Union or in a third country.⁶⁷ This interpretation echoes the product safety approach highlighted above. Some authors go even further, arguing that, because of this rule, European copyright directives "acquire direct effect vis-à-vis private legal entities outside the European Union."⁶⁸

However, there are reasons to be cautious about this interpretation.

First, the provision is contained in a recital of the Al act, not in the enacting terms. Recitals are not legally binding and are meant "to explain the essential objective pursued by the legislative act."⁶⁹ They cannot directly create rights or duties or set norms.⁷⁰

Recital 106 appears intended to aid the interpretation of the provision containing the obligation to which it explicitly refers – namely, Article 53(1)(c) – whereby GPAI providers are required to establish policies to respect copyright and related rights under European law. However, the Recital establishes an additional norm on extraterritoriality that goes beyond or even contradicts the enacting term it references and supports.

The enacting term is Article 53(1)(c). This provision mandates compliance only with European law on copyright and related rights. Based solely on this provision, the logical assumption would be that this applies after conflict of laws rules are considered and within the established framework of copyright territoriality. Applying Recital 106 would mean broadening the territorial scope of the copyright policy obligation and, by extension, substantive copyright rules, particularly the opt-out obligation. However, a non-binding recital of the Al act should not contradict or materially alter the substantive rules of a separate area of law.

To preserve the integrity of the territoriality principle of copyright in this context, some scholars have suggested an "intermediate" solution whereby the policy's obligation in the Al act would only apply "if the model provider scraped websites hosted on servers located in the European Union."⁷¹

This appears to be an attempt to establish a localisation fiction, in which relevant TDM activities are deemed to take place in the European Union, thereby aligning with the *lex loci protectionis* principle.

However, in the context of global internet communications, pinpointing the exact location of servers in the European Union where legally relevant TDM activities occur is not straightforward. This challenge is relevant in European law, since the Court of Justice of the

⁶⁷ Giuseppe B. Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," European Intellectual Property Review, vol. 46, no. 7, 2024, pp. 479–487.

⁶⁸ Malte Stieper and Michael Denga, The International Reach of EU Copyright Through the AI Act (Institut für Wirtschaftsrecht, 2024).

⁶⁹ Maarten den Heijer, Teun van Os van den Abeelen and Antanina Maslyka, "On the Use and Misuse of Recitals in European Union Law," Amsterdam Law School Research Paper No. 2019-31, 2019.

⁷⁰ Maarten den Heijer, Teun van Os van den Abeelen and Antanina Maslyka, "On the Use and Misuse of Recitals in European Union Law," Amsterdam Law School Research Paper No. 2019-31, 2019; European Commission: Legal service, Joint Practical Guide of the European Parliament, the Council and the Commission for Persons Involved in the Drafting of European Union Legislation (Publications Office of the European Union, 2015).

⁷¹ Alexander Peukert, "Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training," pp. 11–12, Social Science Research Network, 2019, identifying problems with the "lex scraping" approach.



European Union (CJEU) requires a minimum level of certainty when determining jurisdiction. For example, in *Wintersteiger*⁷², the Court emphasised that, for the sake of foreseeability, a server of "uncertain location" cannot serve as the basis for determining the place where the event giving rise to the damage occurred. Instead, a "definite and identifiable place" is required.⁷³

Achieving this level of certainty may be especially challenging for TDM. Consider, for instance, how content delivery networks (CDNs) cache content on geographically distributed servers to enhance access speeds. This distribution increases the uncertainty of localising the specific server involved in TDM activities.⁷⁴ Even if it were possible, it would still be necessary to assess the relative significance of the acts localised on that server within the overall TDM activities in question. As a result, this intermediate approach – which seeks to preserve a consistent territorial link – may prove impractical.

Another defence of the AI act's extraterritorial effect on copyright relies on international and EU copyright law, including CJEU case law on infringement localisation. The argument here is that if TDM-related extraction and reproduction are "functionally" essential to AI model training, and those models are made available in the European Union, applying European law is justified, as these acts are part of a broader process linked to the European Union.⁷⁵

The problem with this argument is the clear factual and legal distinction between (1) TDM activities used to train and build a model and (2) making that model available on the European market.⁷⁶ European copyright law treats these separately, as TDM relates to reproduction and extraction rights, while making a trained GPAI model available follows a different legal regime.⁷⁷ Similarly, the AI act differentiates between training a GPAI model and its later integration into an AI system, and different scenarios of commercialisation.⁷⁸ This pro-extraterritoriality argument should therefore be rejected.

To summarise:

- If TDM occurs outside the European Union, GPAI model providers are not required to ensure compliance with Article 4 of the CDSM directive.
- A violation of Recital 106 of the Al act does not amount to copyright infringement but rather a breach of the Al act, leading to administrative penalties rather than copyright enforcement.
- Since the Al act only requires providers to establish "policies to respect copyright," it would be unreasonable to penalise a company for violating European copyright law if it has already complied with the applicable legal framework, including conflict of laws rules.

77 Malte Stieper and Michael Denga, The international reach of EU copyright through the AI Act (Institut für Wirtschaftsrecht, 2024), on "The fact that the model trained abroad is later offered in an EU Member State does not establish a sufficient domestic connection of this act of reproduction."

78 Cf. inter alia the definitions in Art. 3(9)-(11), (63), (66) AI Act.

⁷² Case C-523/10 Wintersteiger ECLI:EU:C:2012:220, 36.

⁷³ Case C-523/10 Wintersteiger ECLI:EU:C:2012:220, 37.

⁷⁴ Michael Veale and João Pedro Quintais, "The Obligations of Providers of General-Purpose AI Models," The Artificial Intelligence Act – A Thematic Commentary, ed. by Gabriela Zanfir Fortuna and others (Hart, 2025).

⁷⁵ Eleanora Rosati, "Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law," European Journal of Risk Regulation, 2024.

⁷⁶ See e.g., Malte Stieper and Michael Denga, The international reach of EU copyright through the AI Act (Institut für Wirtschaftsrecht, 2024) on treating training, design and output as "separate regulatory subjects."

• From a policy standpoint, it is questionable whether a non-binding recital in a law that only indirectly relates to copyright should be used to override fundamental European copyright principles.

Code of Practice and Meta-regulation: A Brief Reference

One possible solution to address the extraterritoriality issue could be to tackle it in the GPAI code of practice. Under Article 53(4) of the AI act, compliance with this code is one of several recognised means for demonstrating compliance with key copyright-related obligations.

The preferred option for compliance is the development of a harmonised European standard in relation to GPAI models.⁷⁹ Compliance with such a standard – covering obligations like transparency and opt-outs – would grant providers a presumption of conformity. Until this standard is published, GPAI model providers can rely on codes of practice to demonstrate compliance.⁸⁰ Article 56 of the AI act regulates these codes, with the AI Office acting as a facilitator. The AI Office and the Board must ensure that codes of practice include at least certain minimum obligations, notably detailing the content used for training.⁸¹

The code aims to clarify key obligations of the Al act for GPAI model providers, including those with systemic risks, and was planned to take effect 12 months after the Al act entered into force. Although these codes are generally "soft law," the Commission may approve them via an implementing act, granting them validity across the European Union.⁸² Bygrave and Schmidt see codes of practice as "embedded meta-regulation" and note that in the Al act, these codes "are probably compulsory..., at least as a 'stop-gap' measure prior to the adoption of harmonised standards."⁸³

The code of practice is currently in its third draft version. The third draft attempts to simplify and clarify the copyright-related obligations of signatories under the Al act. It restructures earlier measures into six main measures (Measures I.2.1 to I.2.6) addressing copyright compliance policies, content-mining practices and mechanisms to prevent infringing Al outputs. The copyright section integrates key performance indicators (KPI) elements (like transparency and complaint procedures) into the measures but no longer lists them separately. It also affirms that these commitments do not override existing European copyright law or private licensing agreements.

Previous versions of the code either mentioned Recital 106 or more clearly suggested a type of voluntary extraterritoriality.⁸⁴ The current version of code scales back these suggestions and does not mandate extraterritoriality or even mention Recital 106. This shift raises the question of whether some of the code's measures can or should be understood as promoting a voluntary and mild approach to extending the copyright policies obligation in Article 53(1)(c) beyond the territorial borders of European law.

⁷⁹ See generally Josep Soler Garrido and others, Harmonised Standards for the European AI Act (Seville: European Commission, 2024).

⁸⁰ A third compliance option applies if no harmonised standard exists and a GPAI model provider does not adhere to an approved code of practice. In this case, the provider must demonstrate alternative adequate compliance measures for Commission approval under Art. 53(3) AI Act. For these and additional compliance options, see Michael Veale and João Pedro Quintais, "The Obligations of Providers of General-Purpose AI Models" in *The Artificial Intelligence Act – A Thematic Commentary*, ed. by Gabriela Zanfir Fortuna and others (Hart, 2025).

⁸¹ Under Art. 56(2) AI Act.

⁸² Lee A. Bygrave and Rebecca Schmidt, "Regulating Non-High-Risk AI Systems under the EU's Artificial Intelligence Act, with Special Focus on the Role of Soft Law," University of Oslo Faculty of Law Research Paper, No. 2024-10. According to Art. 56(9) AI act, since the Code was not ready by 02 May 2025, the Commission may impose common rules for, inter alia, copyright-related obligations under Art. 53.

⁸³ Lee A. Bygrave and Rebecca Schmidt, "Regulating Non-High-Risk AI Systems under the EU's Artificial Intelligence Act, with Special Focus on the Role of Soft Law," University of Oslo Faculty of Law Research Paper, No. 2024-10.

⁸⁴ João Pedro Quintais, "Generative AI, Copyright and the AI Act," Computer Law & Security Review, vol. 56, 2025, p. 106107; Sévérine Dusollier and others, "Copyright and Generative AI," JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law, vol. 16, no. 1, 2025.





The answer to that question is unclear and depends on how the code is interpreted.

On the one hand, it is possible to read several measures (Measures I.2.2–I.2.4) as indicating that signatories voluntarily commit to making "reasonable efforts" to comply with the Al act's copyright policy obligations – including lawful access and opt-out requirements – as they pertain to web crawling activities in model development, regardless of whether (i) these activities occur inside or outside the European Union, and (ii) they are carried out by the signatories themselves or by upstream providers involved in web-crawling activities.

On the other hand, it is also plausible to argue that, since nothing in these measures explicitly references extraterritorial application, they should be interpreted in accordance with European law. That is to say, based on the arguments above, these measures should not apply to TDM activities taking place outside the European Union, whether carried out by the GPAI model provider or related third parties.

The definitive answer will ultimately depend on the final text of the code, particularly on whether it includes sufficiently clear and enforceable commitments, whether key GPAI model providers endorse it, and how its implementation and enforcement are carried out in practice.

Conclusion

This paper has examined the interaction between the AI act and European copyright law, with a particular focus on the contested issue of extraterritoriality. The main findings are set out below.

European copyright law is grounded in the principle of territoriality, meaning it applies to acts that occur and can be localised within the borders of the European Union. TDM activities conducted outside the European Union are governed by the laws of the place where they occur and are not automatically subject to European copyright obligations.

While Recital 106 of the Al act appears to extend copyright compliance obligations extraterritorially, such a provision lacks binding force and contradicts established principles of international and European copyright law. Attempts to impose European copyright standards on non-European TDM activities based on a recital raise legal uncertainties and risk exceeding the intended scope of the legislation. They also do not offer clear tangible benefits, for instance in improving the remuneration of creators.

The TDM acts involved in training a model and the act of placing that model on the European market are legally distinct. On the basis of current European law, compliance with European copyright law should not retroactively be imposed on – or extended to – training conducted lawfully in third countries.

Amid increasing calls to simplify European regulation, particularly in the digital sector, and growing concerns that "inconsistent and restrictive regulation" may hinder technological innovation,⁸⁵ there is even discussion of pausing the Al act to allow for simplification.⁸⁶ In this context, it is essential to minimise the risks of regulatory overreach and legal uncertainty. The complexity of the current framework – particularly the interplay between European copyright law and the Al act, including unresolved issues of extraterritoriality – does little to foster a genuine level playing field. Moreover, the existing rules are unlikely to be fit for the purpose of ensuring fair remuneration for creators in the context of generative Al. Aligning the Al act with the substantive and conflict-of-law rules of European copyright law would enhance legal certainty, facilitate compliance, and more effectively advance the objective of the European Union of promoting Al development.

⁸⁵ Mario Draghi, The Future of European Competitiveness (Brussels: European Commission, 2024).

⁸⁶ Luca Bertuzzi, EU Commission eyes pausing AI Act's entry into application, May 26, *MLex*, <u>https://www.mlex.com/mlex/articles/2344845/eu-commission-eyes-pausing-ai-act-s-entry-into-application</u>.



References

The below list of references excludes legislative instruments, policy documents, and judgements.

99designs Team. "Freelance Design in the Age of Al," 99designs, 22 October 2024

- Abbamonte, Giuseppe B. "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 479–487
- Bertuzzi, Luca. EU Commission eyes pausing Al Act's entry into application, May 26, *MLex*, <u>https://www.mlex.com/mlex/articles/2344845/eu-commission-eyes-pausing-ai-act-s-entry-into-application</u>
- Bulayenko, Oleksandr and others. "Al Music Outputs: Challenges to the Copyright Legal Framework," *Social Science Research Network*, 28 February 2022
- Bygrave, Lee A. and Rebecca Schmidt. "Regulating Non-High-Risk AI Systems under the EU's Artificial Intelligence Act, with Special Focus on the Role of Soft Law," *University of Oslo Faculty of Law Research Paper No. 2024-10*, 24 October 2024
- Chat GPT Is Eating the World (2024)
- Cifrodelli, Gabriele and Lilian Edwards. "Copyright and Generative AI: What Can We Learn from Model Terms and Conditions?," *Kluwer Copyright Blog*, 24 July 2024
- CISAC and PMP Strategy. Study on the Economic Impact of Generative AI in Music and Audiovisual Industries (CISAC, November 2024)
- Cooper, A. Feder and others. "Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models," *arXiv*, 18 May 2025
- Cooper, A. Feder and James Grimmelmann. "The Files Are in the Computer: On Copyright, Memorization, and Generative AI," *arXiv*, 18 July 2024
- Crawford, Lisa Burton. "The Problem of Complex Legislation," Legal Theory, 2024, pp. 1–21
- De Cock, Caroline. "From Upload to Output Filters? How the Al Act's Code of Practice Could Threaten Freedom of Expression and Thought," *Tech Policy Press*, 2025
- Dornis, Tim W. "The Training of Generative AI is Not Text and Data Mining," *European Intellectual Property Review*, no. 2 (2025)
- Draghi, Mario. *The Future of European Competitiveness* (Brussels: European Commission, 2024)
- Dusollier, Séverine and others. "Copyright and Generative AI: Opinion," *JIPITEC Journal of Intellectual Property, Information Technology and E-Commerce Law,* vol. 16, no. 1, 2025
- Edwards, Lilian and others. "Private Ordering and Generative AI: What Can We Learn From Model Terms and Conditions?" *CREATe*, 24 May 2024



- van Eechoud, Mireille M. M. "Territoriality and the Quest for a Unitary Copyright Title," *IIC – International Review of Intellectual Property and Competition Law*, vol. 55, no. 1, 2024, pp. 66–88
- Emanuilov, Ivo and Thomas Margoni. "Forget Me Not: Memorisation in Generative Sequence Models Trained on Open Source Licensed Code," *Zenodo*, February 2024
- European Union Intellectual Property Office (EUIPO). Development of Generative Artificial Intelligence from a Copyright Perspective (Alicante: EUIPO, 2025)
- Evans, Benedict. "Generative AI and Intellectual Property," Benedict Evans, 27 August 2023
- Geiger, Christophe and Vincenzo Iaia. "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI," *Computer Law & Security Review* 52 (2024)
- Goldstein, Paul and P. Bernt Hugenholtz. *International Copyright: Principles, Law, and Practice,* 4th ed. (Oxford: Oxford University Press, 2019)
- Goldstein, Paul, Christianne Stuetzle and Susan Bischoff. "Kneschke vs. LAION Landmark Ruling on TDM Exceptions for AI Training Data – Parts 1 and 2," *Kluwer Copyright Blog*, 13 and 14 November 2024
- Guadamuz, Andres. "First Case on AI and Copyright Referred to the CJEU," *TechnoLlama*, 27 May 2025
- Guadamuz, Andrés. "A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs," *GRUR International*, vol. 73, no. 2, 2024, pp. 111–127
- Hamann, Hanjo. "Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and Their (In)Compatibility with Article 4(3) of the Copyright DSM Directive," *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 15, no. 2, 2024
- den Heijer, Maarten, Teun van Os van den Abeelen and Antanina Maslyka. "On the Use and Misuse of Recitals in European Union Law," *Amsterdam Law School Research Paper No.* 2019-31, 2019
- Heinze, Christian and Cara Warmuth. "Intellectual Property and the Brussels Ibis Regulation" in *Research Handbook on the Brussels Ibis Regulation*, pp. 147–171 (Edward Elgar Publishing, 2020)
- Keller, Paul. "Do Al Models Dream of Dolphins in Lake Balaton?," *Kluwer Copyright Blog*, 28 May 2025
- Keller, Paul and Zuzanna Warso. *Defining Best Practices for Opting out of ML Training* (Open Future Foundation, 2023)
- Kur, Annette and Ulf Maunsbach. "Choice of Law and Intellectual Property Rights," *Oslo Law Review*, vol. 6, no. 1, 2019, pp. 43–61
- Leerssen, Paddy. "Embedded GenAl on Social Media: Platform Law Meets Al Law DSA Observatory," *DSA Observatory Blog*, 16 October 2024

- European Commission: Legal service. *Joint Practical Guide of the European Parliament, the Council and the Commission for Persons Involved in the Drafting of European Union Legislation* (Publications Office of the European Union, 2015)
- Margoni, Thomas. "TDM and Generative AI: Lawful Access and Opt-Outs," *Auteurs&Media*, 2024,
- Margoni, Thomas and Martin Kretschmer. "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," *GRUR International*, vol. 71, no. 8, 2022, pp. 685–701
- Mezei, Péter. "A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI," *European Intellectual Property Review*, vol. 46, no. 7, 2024, pp. 461-469
- Narayanan, Arvind and Sayash Kapoor. "Al as Normal Technology," *Knight First Amendment Institute*, 2025
- Peukert, Alexander. "Copyright in the Artificial Intelligence Act A Primer," *GRUR International*, vol. 73, no. 6 (2024), pp. 497–509
- ——. "Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training," Social Science Research Network, 25 July 2024
- Quintais, João Pedro. "Generative AI, Copyright and the AI Act," *Computer Law & Security Review* vol. 56, 2025, p. 106107
- ——. "What Is a 'Research Organisation' and Why It Matters: From Text and Data Mining to Al Research," GRUR International, vol. 74, no. 5, 2025, pp. 397–398
- Rosati, Eleonora. "CJEU Receives First Referral on Chatbots and Copyright," *The IPKat*, 26 May 2025
- ——. "Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law," European Journal of Risk Regulation, 2024
- Sag, Matthew. "Fairness and Fair Use in Generative AI," *Fordham Law Review*, vol. 92, 2024, p. 1887
- Senftleben, Martin, 'Generative AI and Author Remuneration', *IIC International Review of Intellectual Property and Competition Law*, vol. 54, 2023, pp. 1535–1560
- Silva, Nuno Sousa e. "Are Al Models' Weights Protected Databases?," *Kluwer Copyright Blog*, 18 January 2024
- Soler Garrido, Josep and others. *Harmonised Standards for the European AI Act* (Seville: European Commission, 2024)
- Stieper, Malte and Michael Denga. *The International Reach of EU Copyright Through the AI Act* (Institut für Wirtschaftsrecht, 2024)
- Szkalej, Kacper. "The Paradox of Lawful Text and Data Mining? Some Experiences from the Research Sector and Where We (Should) Go from Here," *Social Science Research Network*, 23 October 2024



- Ueno, Tatsuhiro. "Flexible Copyright Exception for 'Non-Enjoyment' Purposes Recent Amendment in Japan and Its Implication," *GRUR International*, vol. 70, no. 2, 2021, pp. 145–152
- United States Copyright Office (USCO). *Copyright and Artificial Intelligence, Part 3: Generative AI Training (Pre-Publication Version)*, pp. 1–113 (USCO, 2025)
- Valk, Etienne and Iris Toepoel. "DPG Media et al vs. HowardsHome A National Ruling on DSM's Press Publishers' Rights and TDM Exceptions," *Kluwer Copyright Blog*, 16 January 2025
- Veale, Michael and João Pedro Quintais. "The Obligations of Providers of General-Purpose Al Models" in *The Artificial Intelligence Act – A Thematic Commentary*, ed. by Gabriela Zanfir Fortuna and others (Hart, 2025) [Forthcoming]

W3C Community Group. TDM Reservation Protocol (TDMRep) (W3C Community Group, 2024)



Copyright © The Lisbon Council 2025



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Licence

The Lisbon Council asbl IPC-Résidence Palace

155 Rue de la Loi 1040 Brussels, Belgium T +32 2 647 9575 www.lisboncouncil.net info@lisboncouncil.net