Computer Law & Security Review 56 (2025) 106107

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Generative AI, copyright and the AI Act

João Pedro Quintais

University of Amsterdam, Institute for Information Law (IViR), Netherlands

ARTICLE INFO

Keywords:
Generative AI
AI Act
AI
Copyright
Text and data mining
Transparency
Remuneration

ABSTRACT

This paper provides a critical analysis of the Artificial Intelligence (AI) Act's implications for the European Union (EU) copyright acquis, aiming to clarify the complex relationship between AI regulation and copyright law while identifying areas of legal ambiguity and gaps that may influence future policymaking. The discussion begins with an overview of fundamental copyright concerns related to generative AI, focusing on issues that arise during the input, model, and output stages, and how these concerns intersect with the text and data mining (TDM) exceptions under the Copyright in the Digital Single Market Directive (CDSMD).

The paper then explores the AI Act's structure and key definitions relevant to copyright law. The core analysis addresses the AI Act's impact on copyright, including the role of TDM in AI model training, the copyright obligations imposed by the Act, requirements for respecting copyright law—particularly TDM opt-outs—and the extraterritorial implications of these provisions. It also examines transparency obligations, compliance mechanisms, and the enforcement framework. The paper further critiques the current regime's inadequacies, particularly concerning the fair remuneration of creators, and evaluates potential improvements such as collective licensing and bargaining. It also assesses legislative reform proposals, such as statutory licensing and AI output levies, and concludes with reflections on future directions for integrating AI governance with copyright protection.

1. Introduction

This paper explores the intersection of the Artificial Intelligence (AI) Act[1] and the EU copyright acquis,[2] focusing on their implications for generative AI and text and data mining (TDM). Its primary aim is to provide a comprehensive analysis – descriptive, critical, and normative – of the relationship between these two frameworks. To facilitate an in-depth examination of the complexity of EU law in this area, the paper does not explore in detail international perspectives, which would merit a separate analysis.[3] By identifying potential gray areas and blind spots within the existing EU legal framework, the paper offers a detailed critique of its legal ambiguities. It further explores the practical and theoretical implications of these issues, aiming to inform both current and future policymaking at the EU level. The intended audience includes policymakers, regulators, and academics engaged in AI and copyright law, particularly within the EU, while also providing valuable insights for practicing attorneys and copyright holders navigating the evolving legal landscape. Ultimately, the paper seeks to advance interpretative proposals that contribute meaningfully to ongoing debates and policy discussions in this rapidly developing area of law.

E-mail address: j.p.quintais@uva.nl.

[1] 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) (AI Act).

[2] This paper builds on a previous analysis of the European Parliament version of these rules: João Pedro Quintais, "Generative AI, Copyright and the AI Act" (*Kluwer Copyright Blog*, May 9, 2023) <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/> accessed August 18, 2023. For more recent analyses, see Paul Keller, "A First Look at the Copyright Relevant Parts in the Final AI Act Compromise" (*Kluwer Copyright Blog*, December 11, 2023) <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/> accessed July 31, 2024; Alexander Peukert, "Copyright in the Artificial Intelligence Act – A Primer" (2024) 73 GRUR International <https://academic-oup-com.proxy.uba.uva.nl/grurint/article/73/6/497/7675073> accessed July 31, 2024; Giuseppe B Abbamonte, "The Application of the Copyright TDM Exceptions and Transparency Requirements in the AI Act to the Training of Generative AI" (2024) 46 European Intellectual Property Review 479.

[3] See however section 2, which identifies relevant scholarship and rules in other jurisdictions facign similar issues.

https://doi.org/10.1016/j.clsr.2025.106107

Available online 29 January 2025

Table 1
Scope EU TDM exceptions.

Provision → Scope & Conditions ↓	Article 3	Article 4
Acts	Text and data mining	Reproductions and extractions
Rights	Database Directive: 5(a) and 7(1) InfoSoc Directive: 2 CDSMD: 15(1)	Database Directive.: 5(a) and 7(1) InfoSoc Directive: 2 Software Directive.: 4(1)(a) and (b) CDSMD: 15(1)
Purpose of use	Scientific research	Text and data mining
Beneficiaries	Research organisations; Cultural heritage institutions	All (unrestricted)
Lawful access	Yes	Yes
Other conditions for beneficiary	Works must be stored with "appropriate level of security"	–
Contractual Derogation	No	Yes
Other Reservations or Exclusions	No	Yes: via "machine-readable means in the case of content made publicly available online"
Other Conditions	Three-step test; Partial technical protection measures protection under 6(4) InfoSoc Directive	Three-step test; Partial technical protection measures protection under 6(4) InfoSoc Dir.
Connection to other exceptions in *acquis*	InfoSoc Directive.: 5(3)(a)	InfoSoc Dirctive: 5(1)

The paper proceeds as follows. After this short introduction, Section 2 outlines the basic copyright issues of generative AI and the relevant copyright *acquis* rules that interface with the AI Act. It mentions potential copyright issues with the input or training stage, the model, and outputs. The AI Act rules are mostly relevant for the training of AI models, and the Regulation primarily interfaces with the TDM exceptions in Articles 3 and 4 of the Copyright in the Digital Single Market Directive (CDSMD).[4] Section 3 then briefly explains the AI Act's structure and core definitions as they pertain to copyright law. Section 4 is the heart of the paper. It covers in some detail the interface between the AI Act and EU copyright law, namely: the clarification that TDM is involved in training AI models (4.1); the outline of the key copyright obligations in the AI Act (4.2); the obligation to put in place policies to respect copyright law, especially regarding TDM opt-outs (4.3); the projected extraterritorial effect of such obligations (4.4); the transparency obligations (4.5); how the AI Act envisions compliance with such obligations (4.6); and potential enforcement and remedies (4.7). Section 5 then provides normative insights on the inadequacy of the current regime to address one of its main concerns – the fair remuneration of authors and performers – and argues for possible avenues to improve the current regime, with an emphasis on collective licensing and collective bargaining. This section also provides a critical analysis of existing proposals for legislative reform, including statutory licensing and AI output levies. The purpose of this examination is not to advance a standalone legal reform proposal but rather to offer interpretative options based on a critical analysis of the existing EU legal framework, thereby contributing to debates on feasible approaches to fair remuneration of creators in this area. Section 6 concludes.

2. Inputs, model, outputs and EU copyright law

In simple terms, it is possible to find copyright questions at different stages of the AI lifecycle or value chain. These questions relate to the input or training stage, the model itself and the outputs generated by (or with the assistance of) an AI model or system.[5]

From the input perspective, training and developing AI models might involve a number of activities (e.g. web scraping, pre-training, training) that often entail copyright relevant reproductions. In EU copyright law, many such activities qualify as TDM and are predominantly regulated by different exclusive rights of reproduction and the two TDM exceptions in the CDSMD.[6]

TDM is defined in Article 2(2) of the directive 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." In EU law, it is now mostly settled that this definition covers lato senso pre-training and training activities needed to develop an AI model, including generative AI models.[7] The AI Act reiterates this view. It is still an open question whether the concept of TDM would cover *all* the copyright-relevant acts that might be involved in training and building an AI model.[8]

Articles 3 and 4 CDSMD contain two TDM-related mandatory exceptions. Article 3 applies to acts of TDM for the purposes of scientific research by "research organisations" and "cultural heritage institutions", regarding works/subject matter to which they have lawful access, and subject to a number of additional conditions.[9]

Article 4 provides an exception for reproductions and extractions of *lawfully accessed* works/subject matter for the purposes of TDM. This exception is subject to reservation by rights holders, including through "machine-readable means in the case of content made publicly available online", for instance through the use of metadata and terms and conditions of a website or a service. Such reservation shall not affect the application of the TDM exception for scientific purposes in Article 3.

[4] 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.) (CDSMD). For general comments on the CDSMD, see João Pedro Quintais, "The New Copyright in the Digital Single Market Directive: A Critical Look" (2020) 42 European Intellectual Property Review 28; Séverine Dusollier, "The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition" (2020) 57 Common Market Law Review 979.

[5] For an analysis of some of the issues described below under EU and UK law, see Andrés Guadamuz, "Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs" (2024) 73 GRUR International 111.

[6] Generally on the CDSMD's TDM exceptions, see Thomas Margoni and Martin Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology" (2022) 71 GRUR International 685; Rossana Ducato and Alain Strowel, "Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out" (2021) 43 European Intellectual Property Review 322.

[7] For an argument that training and AI model should not qualify as TDM, see Tim W Dornis, "The Training of Generative AI Is Not Text and Data Mining" [2025] European Intellectual Property Review <https://papers.ssrn.com/abstract=4993782> accessed January 13, 2025.

[8] See infra at 4.1.

[9] This provision may be combined with the optional exception covering uses for non-commercial scientific research purposes in Article 5(3)(a) Directive 2001/29/EC (InfoSoc Directive), which already covered certain TDM activities. On this interaction see e.g. Margoni and Kretschmer (n 6); Quintais, "The New Copyright in the Digital Single Market Directive" (n 4); Kacper Szkalej and Martin Senftleben, "Generative AI and Creative Commons Licences - The Application of Share Alike Obligations to Trained Models, Curated Datasets and AI Output" (2024) 15 JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law <https://www.jipitec.eu/jipitec/article/view/415> accessed January 13, 2025.

This possibility of rights reservation is commonly referred to as the "opt-out" requirement. Both the lawful access and the opt-out requirements are crucial to understand the AI Act provisions discussed below.[10]

Table 1 below provides a summary overview of the scope of the EU TDM exceptions. ,[11] ,[12][13]

At this stage, most of the action in the EU is taking place at the legislative and policy level. A notable exception is a recent German decision, likely the first EU ruling on the new TDM exceptions with relevance to AI.[14] On 27 September 2024, the District Court of Hamburg published its decision in the case of *Kneschke v LAION*.[15] The case concerned the legality of using copyright-protected works (in this case, a photograph by Kneschke) for training generative AI models. LAION, a non-profit organization, provides datasets, tools, and models for machine learning research. Some of these datasets and tools are subsequently used by commercial generative AI providers, such as Stable Diffusion.

Essentially, the German court sided with LAION, recognizing it as a "research organization" and ruling that its activities were covered by the TDM exception for scientific purposes under the German law transposing Articles 2(1) and 3 of the CDSMD.[16] This ruling is significant not only for its precedential value and some of the reasoning (discussed below) but also because it highlights critical shortcomings in the AI Act's copyright-relevant rules. Notably, the Act does not apply to the TDM activities of entities like LAION unless they also qualify as model providers.

The EU approach is different from that of other jurisdictions.[17] In the US, for instance, without a specific TDM exception, the focus is on how fair use applies to the training and development of AI models that relies on massive amounts of copyrighted works. At time of writing, there are >30 cases before US courts that tackle this question from different angles.[18] The current policy debate in Canada seems to toggle between the EU and US approaches.[19] In Japan, the existence of a "non-enjoyment" exception in the Copyright Act that covers TDM activities would appear to leave significant space for lawful AI training.[20] In Israel, altough the topic is debated and has yet to be decided in court, the Office of Legal Counsel and Legislative Affairs of the Israel Ministry of Justice has for instance issued an opinion to the effect that use of copyrighted materials for machine learning purposes is generally permitted under the joint provisions of fair use, incidental use and transient use in the Israeli Copyright Act.[21] Finally, in Latin America, there is significant uncertainty on the legal status of TDM, particularly as regards research activities, leading to calls for legal reform that take into account the socio-economic and cultural characteristics of the region.[22]

From the AI model perspective, the main legal questions that arise thus far relate to the whether the model weights qualify as protected databases[23] and whether there is sufficient memorization of protected content in a model for the same to qualify as containing (unauthorized) reproductions of that content.[24]

From the output perspective, key questions include whether AI-

[10] See infra at 4, in particular 4.4.

[11] Directive 96/9 on the legal protection of databases [1996] OJ L77/20 (Database Directive).

[12] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc Directive).

[13] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) (Software Directive)

[14] A subsequent first instance decision in the Netherlands is also notable. See ECLI:NL:RBAMS:2024:6563 - Rechtbank Amsterdam, 30-10-2024 / C/13/737170 / HA ZA 23-690.

[15] Case No. 310 O 227/23, https://drive.google.com/file/d/1A_vnSJUwlrVovhIqA4rKIFaOktR4TvBt/view

[16] For early comment on this decision, see Eleonora Rosati, "The German LAION Decision: A Problematic Understanding of the Scope of the TDM Copyright Exceptions and the Transition from TDM to AI Training" (*The IPKat*, 2024) <https://ipkitten.blogspot.com/2024/10/the-german-laion-decision-problematic.html> accessed October 10, 2024; Paul Keller, "LAION vs Kneschke: Building Public Datasets Is Covered by the TDM Exception" (*Open Future*, 2024) <https://openfuture.eu/blog/laion-vs-kneschke> accessed October 10, 2024; Jonathan Pukas and Jan Bernd Nordemann, "German Regional Court (Landgericht) of Hamburg Paves the Way to Treat the Reproduction of Works as AI Training Data under the EU Text and Data Mining Exceptions" (*Kluwer Copyright Blog*, October 25, 2024) <https://copyrightblog.kluweriplaw.com/2024/10/25/german-regional-court-landgericht-of-hamburg-paves-the-way-to-treat-the-reproduction-of-works-as-ai-training-data-under-the-eu-text-and-data-mining-exceptions/> accessed October 25, 2024; 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*, November 13, 2024) <https://copyrightblog.kluweriplaw.com/2024/11/13/kneschke-vs-laion-landmark-ruling-on-tdm-exceptions-for-ai-training-data-part-1/> accessed January 13, 2025.

[17] For overviews of different jurisdiction and their rules on TDM, see e.g. Sean M Fiil-Flynn and others, "Legal Reform to Enhance Global Text and Data Mining Research" (2022) 378 Science 951; Alexander Peukert, "Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training" (Faculty of Law of Goethe University Frankfurt/M 2024) Research Paper of the Faculty of Law of Goethe University Frankfurt/M 3/2024 <https://papers.ssrn.com/abstract=4905400> accessed January 27, 2025; Matthew Sag and Peter K Yu, "The Globalization of Copyright Exceptions for AI Training" (2025) 74 Emory Law Journal <https://papers.ssrn.com/abstract=4976393> accessed January 27, 2025.

[18] See "Chat GPT Is Eating the World" (*Chat GPT Is Eating the World*, 2024) <https://chatgptiseatingtheworld.com/> accessed January 27, 2025. for a running list of generative AI lawsuits int he US. It is impossible to list here all US scholarship on this topic, but for an in-depth analysis referring to much of this scholarship, see e.g. Katherine Lee, A Feder Cooper and James Grimmelmann, "Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain" (arXiv, March 1, 2024) <http://arxiv.org/abs/2309.08133> accessed January 27, 2025; Matthew Sag, "Copyright Safety for Generative AI" (2023) 61(2) Houston Law Review <https://papers.ssrn.com/abstract=4438593> accessed January 27, 2025.

[19] Carys J Craig, "Canada's Changing AI-Copyright Policy Discourse: A Play in Three Parts?" (*Kluwer Copyright Blog*, April 25, 2024) <https://copyrightblog.kluweriplaw.com/2024/04/25/canadas-changing-ai-copyright-policy-discourse-a-play-in-three-parts/> accessed July 31, 2024; Carys J Craig, "The AI-Copyright Trap" (Social Science Research Network, August 28, 2024) <https://papers.ssrn.com/abstract=4905118> accessed January 27, 2025.

[20] Tatsuhiro Ueno, "Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication | GRUR International | Oxford Academic" (2021) 70 GRUR International 145.

[21] Jonathan Band, "Israel Ministry of Justice Issues Opinion Supporting the Use of Copyrighted Works for Machine Learning" (*Disco - Disruptive Competition Project*, January 19, 2023) <https://project-disco.org/intellectual-property/011823-israel-ministry-of-justice-issues-opinion-supporting-the-use-of-copyrighted-works-for-machine-learning/> accessed January 13, 2025. The opinion mentioned is available at https://www.gov.il/BlobFolder/legalinfo/machine-learning/he/machine-learning.pdf

[22] Luca Schirru and others, "Text and Data Mining Exceptions in Latin America" [2024] IIC - International Review of Intellectual Property and Competition Law <https://doi.org/10.1007/s40319-024-01511-2> accessed January 27, 2025.

[23] Nuno Sousa e Silva, "Are AI Models' Weights Protected Databases?" (*Kluwer Copyright Blog*, January 18, 2024) <https://copyrightblog.kluweriplaw.com/2024/01/18/are-ai-models-weights-protected-databases/> accessed January 27, 2025.

[24] A Feder Cooper and James Grimmelmann, "The Files Are in the Computer: On Copyright, Memorization, and Generative AI" (arXiv, July 18, 2024) <http://arxiv.org/abs/2404.12590> accessed January 27, 2025; Ivo Emanuilov and Thomas Margoni, "Forget Me Not: Memorisation in Generative Sequence Models Trained on Open Source Licensed Code" (Zenodo, February 2024) <https://zenodo.org/records/10635479/preview/GenAI_Memorisation_Open_Source_IEmanuilov_TMargoni-Preprint_Zenodo.pdf?include_deleted=0> accessed January 27, 2025; Julio Carvalho, "The Stubborn Memory of Generative AI: Overfitting, Fair Use, and the AI Act" (*Kluwer Copyright Blog*, April 8, 2024) <https://copyrightblog.kluweriplaw.com/2024/04/08/the-stubborn-memory-of-generative-ai-overfitting-fair-use-and-the-ai-act/> accessed January 27, 2025.

generated outputs are protected by copyright,[25] whether outputs are derivative works, and if they infringe on third-party works used in training.[26] The latter assessment is often bundled with the memorization issue but, as Cooper and Grimmelman note, should in fact be considered in connection with *regurgitation, extraction*, and *reconstruction* at the output generation stage.[27]

In the EU, an increasingly interesting – and so far under-researched – question is whether and to what extent copyright exceptions apply to these AI-generated outputs. This would include for instance freedom of expression "transformative" use or content exceptions like those for: quotation, criticism, review; or use for the purpose of caricature, parody or pastiche.[28] In the EU, it will be particularly relevant to see whether and how the upcoming ruling on *Pelham II*[29] will qualify and define "pastiche" as an autonomous concept of EU law, which could potentially apply to AI-generated outputs.[30]

Related to the issues above, there are also significant legal questions surrounding the relationship between copyright and private ordering of AI outputs by AI model and system providers in the terms and conditions of their services.[31]

Finally, probably the most pressing overarching policy concern in this area is the extent to which creators can and should be remunerated by the use of their works in generative AI models and systems[32]. I return to this topic below in Section 5, once the nuances of the legal regimes and interactions between EU copyright law and the AI Act are clear, including the potential triggers for remuneration.

3. The AI Act: structure and key definitions

The AI Act was published as Regulation 2024/1689 of 13 June 2024 (and in the Official Journal on 12 July).[33] The Act is an extremely long and complex legislative text.[34] Even for a highly motivated academic, it is a challenging and tedious read. The AI Act contains 180 recitals, 113 articles and 13 annexes. Structurally, it is divided into 13 chapters: General Provisions (I), Prohibited AI Practices (II), High-Risk AI Systems (III), Transparency Obligations for Providers and Deployers of Certain AI Systems (IV), General-purpose AI Models (V), Measures in Support of Innovation (VI), Governance (VII), EU Database for High-Risk AI Systems (VIII), Post-Market Monitoring, Information Sharing and Market Surveillance (IX), Codes of Conduct and Guidelines (X), Delegation of Power and Committee Procedure (XI), Penalties (XII), and Final Provisions (XIII).

From the copyright perspective, the most relevant provisions are found in Chapter V, on general purpose AI models, which contains the AI Act's copyright-relevant obligations.[35] According to Article 113, these copyright obligations will enter into force on 2 August 2025.

To understand the copyright issues, it is important to have a basic grasp of key definitions. Specifically, we need to understand the differences between AI systems (Article 3(1)), AI models (undefined), general purpose AI (GPAI) *systems* (Article 3(66)) and GPAI models (Article 3 (63)). For our purposes, it is sufficient to make the following points.

First, as explained in Recital 97, an AI system is different from a GPAI model. GPAI models are characterized by their versatility and ability to perform various tasks. They are usually trained on large datasets using methods like self-supervised, unsupervised, or reinforcement learning. GPAI models can be distributed in multiple formats, such as application programming interfaces (APIs) or direct download, and can be modified or integrated into AI systems, which require additional components like

[25] See, e.g. P Bernt Hugenholtz and João Pedro Quintais, "Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?" [2021] IIC - International Review of Intellectual Property and Competition Law <https://doi.org/10.1007/s40319-021-01115-0> accessed January 27, 2025; Ole Andreas Rognstad, "Creations Caused by Humans (or Robots)? Artificial Intelligence and Causation Requirements for Copyright Protection in EU Law" in Taina Pihlajarinne and Annette Alén-Savikko (eds), *Artificial Intelligence and the Media - Reconsidering Rights and Responsibilities* (Edward Elgar Publishing 2022) <https://papers.ssrn.com/abstract=4843280> accessed January 27, 2025.

[26] On the issue of liability for the generation of potentially infringing outputs see Eleonora Rosati, "Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law" [2024] European Journal of Risk Regulation 1–25, doi:10.1017/err.2024.72, accessed January 26, 27, 2025; Z Krokida, "Large Language Models and EU Intermediary Copyright Liability: Quo Vadis?" (2024) 46 European Intellectual Property Review 361.

[27] Cooper and Grimmelmann (n 24).

[28] See Articles 5(3)(d) and (k) InfoSoc Directive and 17(7) CDSMD.

[29] Request for a preliminary ruling from the Bundesgerichtshof (Germany) lodged on 25 September 2023 – CG and YN v Pelham GmbH and Others (Case C-590/23, Pelham) (*Pelham II*). On which, see European Copyright Society, "Opinion of the European Copyright Society on CG and YN v Pelham GmbH and Others, Case C-590/23 (Pelham II)" (European Copyright Society 2024) ECS Opinion.

[30] On recent developments regarding pastiche, see e.g. Péter Mezei and others, "Oops, I Sampled Again … the Meaning of 'Pastiche' as an Autonomous Concept Under EU Copyright Law" [2024] IIC - International Review of Intellectual Property and Competition Law <https://doi.org/10.1007/s40319-024-01495-z> accessed January 27, 2025; Piero Casanova, "Permissible Pastiche in Pelham II: A Proposed Response" (*Kluwer Copyright Blog*, April 11, 2024) <https://copyrightblog.kluweriplaw.com/2024/04/11/permissible-pastiche-in-pelham-ii-a-proposed-response/> accessed January 27, 2025; Sabine Jacques, "The Parody Exception: Revisiting the Case for a Distinct Pastiche Exception" (*Kluwer Copyright Blog*, October 5, 2023) <https://copyrightblog.kluweriplaw.com/2023/10/05/the-parody-exception-revisiting-the-case-for-a-distinct-pastiche-exception/> accessed January 27, 2025; Susan Bischoff, "The Dawn of Pastiche: First Decision on New German Copyright Exception" (*Kluwer Copyright Blog*, June 7, 2023) <https://copyrightblog.kluweriplaw.com/2023/06/07/the-dawn-of-pastiche-first-decision-on-new-german-copyright-exception/> accessed January 27, 2025; Till Kreutzer, "The Pastiche in Copyrigth Law. Expert Opinion on Pastiche: More Freedom for Social Communication and Internet Culture in Copyright Law" (Gesellschaft für Freiheitsrechte / Rechtsgutachten: Politische Betätigung gemeinnütziger Körperschaften2 Gesellschaft für Freiheitsrechte eV 2023) Expert Opinion <https://freiheitsrechte.org/en/themen/demokratie/expert-opinion-on-pastiche> accessed January 27, 2025; Söğüt Atilla, "ECS's Opinion on Pelham II and Its Potential Implications for AI-Generated Pastiches – Part 2" (*The IPKat*, January 10, 2025) <https://ipkitten.blogspot.com/2025/01/ecss-opinion-on-pelham-ii-and-its.html> accessed January 27, 2025.

[31] See, e.g. Lilian Edwards and others, "Private Ordering and Generative AI: What Can We Learn From Model Terms and Conditions?" (CREATe 2024) <https://zenodo.org/records/11276105> accessed January 27, 2025; Gabriele Cifrodelli and Lilian Edwards, "Copyright and Generative AI: What Can We Learn from Model Terms and Conditions?" (*Kluwer Copyright Blog*, July 24, 2024) <https://copyrightblog.kluweriplaw.com/2024/07/24/copyright-and-generative-ai-what-can-we-learn-from-model-terms-and-conditions/> accessed January 27, 2025. For a US perspective, see Mark A Lemley and Peter Henderson, "The Mirage of Artificial Intelligence Terms of Use Restrictions" (Princeton University Program in Law & Public Affairs Research Paper No 2025-04, December 9, 2024) <https://papers.ssrn.com/abstract=5049562> accessed January 27, 2025.

[32] On which topic, see e.g. Martin Senftleben, "Generative AI and Author Remuneration" (2023) 54 IIC - International Review of Intellectual Property and Competition Law 1535; Christophe Geiger and Vincenzo Iaia, "The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI" (2024) 52 Computer Law & Security Review 105925.

[33] For some of the legislative story of the copyright provisions in the AI Act, see Quintais, "Generative AI, Copyright and the AI Act" (n 2); Peukert (n 2).

[34] On the larger debate of complex legislation and its perils, see Lisa Burton Crawford, "The Problem of Complex Legislation" [2024] Legal Theory 1.

[35] See also supporting recitals 104 to 109

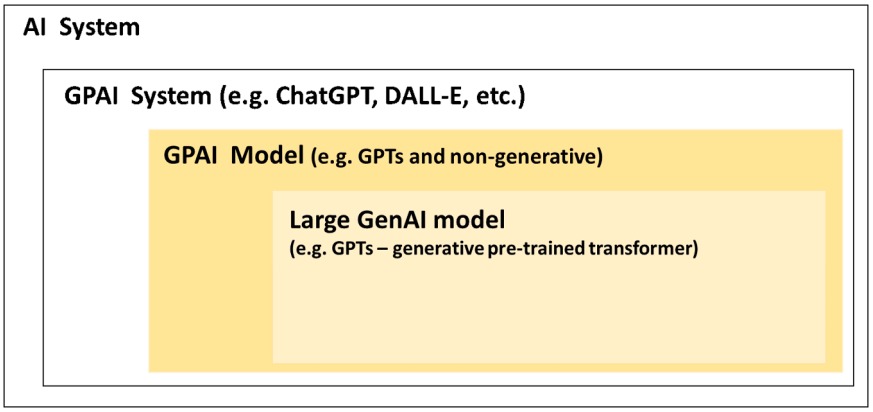

Fig. 1. Representation of AI models and systems.

user interfaces. The AI Act outlines specific rules for GPAI models, particularly those posing systemic risks, ensuring these rules apply when models are placed on the market. The obligations also cover models integrated into AI systems, but exempt models used solely for internal processes or research before market release.

In a nutshell, GPAI models are part of (GP)AI systems. Using a simple analogy, the model is the engine, whereas the system is the car one drives on the road. An example of the model is the generative pre-trained transformers or GPTs; an example of the system is the chatbot or tool many use on a daily basis, such as ChatGPT, Midjourney, Dall-E, or Firefly.[36] As we will see below, this distinction matters for the application and enforcement of the AI Acts's copyright rules.

Second, although the AI Act does not define generative AI, it does clarify in recitals 99 and 105 that large generative AI models are typical *examples* of GPAI models, since they allow for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks. This also means that other types of GPAI models are regulated beyond those that are of the generative type, e.g. those that are included in recommender systems.

Fig. 1 below is a simplified representation of how these definitions relate to each other. For now, it is worth noting that most copyright obligations in the AI Act are imposed on *GPAI model providers* in relation to all types of GPAI models, rather than on GPAI *system providers* or in relation to any type of AI system. Note that in the logic of the AI Act, a *provider* is the party that places a model or system on the market whereas a *deployer* is that which uses an AI system under its authority.[37] Importantly, entities that are involved in distinct pre-training and training activities (e.g. website scraping and dataset creation) that do not include the development and provision of the model itself are not GPAI model providers, even if they are carrying out TDM activities. An example is LAION, in those cases where it provides tools and datasets.

The most relevant scope exclusions here are stated in Articles 2(6) and 2(8).[38] Article 2(6) clarifies that the AI Act does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development. Nevertheless, developers of these models are voluntarily encouraged to comply with the legal requirements. Article 2(8) excludes the application of the Act to "any research, testing, or development activity regarding AI systems or AI models prior to their being placed on the market or put into service." These exclusions are relevant for copyright purposes because if a research organization or cultural heritage institution is doing TDM for scientific purposes and benefits from the exception in Article 3 CDSMD, then it will likely not have to comply with the additional obligations in the AI Act. Crucially, however, the exclusion in Article 2(6) does not apply to: the modifications made to a scientific research model that is then commercialized[39]; or models based on scientific purpose/use GPAI models that are subsequently made available on the market.[40]

Arguably, these exclusions may also benefit entities like LAION, even in scenarios where they act as model providers, as long as their activities are for scientific research purposes. A crucial aspect of this assessment is the fact that, because these entities do not release GPAI models in the context of a commercial activity, they are not "making available on the market" that model under the terms of Articles 3(9) and (10) AI Act.

4. The AI Act vs EU copyright law interface

The AI Act does not mix well with copyright law. The AI Act is conceptually akin to a public law instrument designed through a product safety prism; it is primarily aimed at serving the public interest[41] through the imposition of systemic compliance obligations on certain providers.[42] Differently, copyright law is mostly an area of private law that affords private rights holders a legal entitlement (considered a fundamental right under Article 17(2) Charter), in the form of an exclusive right or, less frequently, a remuneration right or claim. The public vs private differences in nature lead to differences in enforcement and remedies.[43] As Peukert puts it, the copyright provisions of the AI Act are "a fusion of two different types of laws".[44]

In addition, like with the Digital Services Act (DSA) ,[45] the

[36] See, respectively, https://chatgpt.com/, https://www.midjourney.com/home, https://openai.com/product/dall-e-2, and https://www.adobe.com/products/firefly.html .

[37] Article 3(3) and (4)) AI Act.

[38] See also recitals 25 and 109 AI Act.

[39] Recital 109 AI Act.

[40] Recital 25 AI Act.

[41] See e.g. Article 1(1) AI Act.

[42] Generally on the AI Act, see e.g. Nuno Sousa e Silva, "The Artificial Intelligence Act: Critical Overview" (Social Science Research Network, July 30, 2024) <https://papers.ssrn.com/abstract=4937150> accessed October 31, 2024; Sandra Wachter, "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond" (2024) 26 Yale Journal of Law & Technology 671.

[43] See infra at 4.6 and 4.7.

[44] Peukert (n 2).

[45] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) (DSA). On the interactions between the copyright acquis and the DSA see See João Pedro Quintais and Sebastian Felix Schwemer, "The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright?" (2022) 13 European Journal of Risk Regulation 191; Alexander Peukert and others, "European Copyright Society – Comment on Copyright and the Digital Services Act Proposal" (2022) 53 IIC - International Review of Intellectual Property and Competition Law 358.

interaction between the AI Act and EU copyright law is far from crystal clear. Recital 108 states that the AI Act "does not affect the enforcement of copyright rules as provided for under Union law" and recital 109 mentions that compliance with AI Act obligations is "without prejudice to Union copyright law". But the exclusions in scope in Article 2 AI Act do not explicitly mention copyright law, as they do for instance the topic of liability of intermediaries under Chapter II of the DSA. The complexity is amplified by the fact that the relevant EU copyright law provisions that interface with the AI Act – the TDM exceptions – are part of the CDSMD, meaning that they are susceptible to (slightly to moderately) different implementations in the national laws of 27 member states.

4.1. TDM and copyright

The AI Act takes a clear position on the copyright relevant nature of TDM. Recital 105 states that generative AI models "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." It adds that TDM techniques used to develop these models may require vast amounts of copyright-protected materials, which use "requires the authorisation of the rights holder concerned unless relevant copyright exceptions and limitations apply", namely those in Articles 3 and 4 CDSMD. Recital 105 goes on to mention the rights reservation mechanism in Article 4(3), noting that where applicable GPAI model providers must abide by this mechanism if they want to carry out TDM on those materials.

In essence, Recital 105 is another nail in the coffin of arguments for the non-expressive nature of TDM reproductions of protected works in the context of generative AI models, and thus their lack of relevance for copyright purposes.[46] In the EU at least, it appears that if you do TDM on copyright protected content, you are doing a reproduction of a work. As such, in order to do it lawfully, you either get authorization from the rights holder or benefit from a copyright exception, e.g. in Articles 3 and 4 CDSMD.

But does this mean that all copyright-relevant reproductions and extractions involved in the training and development of an AI model qualify as TDM? Some authors, like Abbamonte, conclude from the applicable framework that is indeed the case.[47] To be sure, the definition of TDM in Article 2(2) CDSMD and supporting recitals is quite broad, and Recital 105 of the AI Act suggests that TDM techniques are crucial for the development of these models. Other authors, like Rosati, argue that such conclusion cannot be derived from the AI Act and note that subsequent acts of reproduction related to dataset search and retrieval (e.g., for the LAION dataset) may fall outside the scope of the TDM exception.[48] Some authors still consider the training generative AI model as excluded from the scope of the TDM definition,[49] a position that is difficult to square with the EU legal provisions discussed here.

What appears clear is that the TDM exception does not cover subsequent acts of communication to the public or making available of the TDM results. Indeed, as highlighted in Section 2 above, the scope of the TDM exceptions covers only acts of reproduction and extraction. Furthermore, Articles 3(2) and 4(2) put clear boundaries on the subsequent uses of copies of works or other subject matter made pursuant to the TDM exceptions.

One illustration of this type of use is the making available to the public of a dataset resulting from the TDM activities.[50] Naturally, this act only has copyright relevance to the extent that the dataset reproduces protected works or subject matter, or where its availability otherwise infringes on a restricted act. The LAION dataset of image-text pairs at issue in *Kneschke v LAION* is a particularly interesting case since it mostly consists of hyperlinks to publicly accessible images.[51] Hyperlinks are not protected works but may qualify as an act of communication to the public under Article 3 InfoSoc Directive. According to established case law of the CJEU, posting hyperlinks to protected works is permissible only if the content linked to is freely available online with the permission of the rightsholders and does not circumvent technical restrictions (e.g., paywalls) and, arguably, some level of contractual restrictions.[52]

4.2. Key copyright obligations in articles 53(1) AI Act and shared regime

The AI Act then establishes key copyright related obligations for all GPAI *model* providers in Article 53(1)(c) and (d).

- First, these providers must put in place a policy to respect EU Union copyright law in particular to identify and respect, including through state-of-the-art technologies, the reservations of rights (i.e. "opt-out") expressed pursuant to Article 4(3) CDSMD.
- Second, these providers must draw up and make publicly available a sufficiently detailed summary about the content used for training of the GPAI model, according to a template provided by the AI Office.

Before discussing each obligation, some shared common features should be highlighted.[53]

First, because the obligations are imposed on GPAI *model* providers, it is not clear that they reach other players in the AI value chain, such as providers of entities responsible for web scraping and crawling (e.g. Common Crawl[54]), providers of training tools and datasets (e.g. LAION) and – perhaps most importantly – AI *system* providers. To be sure, regarding, the latter, the two (model and system) providers may be the same entity, as currently seems to be the case for the bigger players.

According to Recital 97, when the GPAI model provider "integrates an own model into its own AI system that is made available on the market or put into service, that model should be considered to be placed on the market and, therefore, the obligations in this Regulation for models should continue to apply in addition to those for AI systems."[55] But what happens if a model provider does not vertically integrate the model in a system? Does the subsequent (GP)AI system provider – a "downstream provider" under Article 3(68) – have to comply with these obligations? As Peukert notes, and I agree, compliance in this scenario "requires access to and control over the model, which is only available to the model provider".[56] This remains true despite the obligation to draw

[46] See, e.g. Martin Senftleben, "Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?" [2022] IIC - International Review of Intellectual Property and Competition Law <https://link.springer.com/article/10.1007/s40319-022-01266-8> accessed January 27, 2025.

[47] Abbamonte (n 2) 480.

[48] Rosati, "Infringing AI" (n 26).

[49] Dornis (n 7).

[50] Rosati, "Infringing AI" (n 26).

[51] LAION's training data set essentially contains only URLs to the photos it links to, together with the associated image descriptions and related metadata.

[52] João Pedro Quintais, "Untangling the Hyperlinking Web: In Search of the Online Right of Communication to the Public" (2018) 21 The Journal of World Intellectual Property 385; Eleonora Rosati, "Linking and Copyright in the Shade of VG Bild-Kunst" (2021) 2021 Common Market Law Review 1875.

[53] I do not discuss here issues of temporal application, on which see e.g. Rosati, "Infringing AI" (n 26); Peukert (n 2). In essence, the principle of non-retroactivity would dictate that GPAI models made available on the EU market before the entry into force of the AI Act would not be subject to its obligations. Where a model is subsequently subject to modification and fine-tuning and then made available on the EU market after the entry to force of the Act, then it is arguable that its obligations (including those related to copyright) may apply, at least to that modification or fine-tuning. See also Articles 111 and 113 AI Act on temporal application to GPAI models already available on the market.

[54] Common Crawl, https://commoncrawl.org/

[55] For the definitions of "making available on the market" and putting into seervice", see Article 3(10) and (11) AI Act.

[56] Peukert (n 2).

up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the GPAI model into their AI systems.[57]

As AI technology and platforms converge, we can take this analysis one step further. If the model is integrated in a system that is itself embedded in a very large online platform (VLOP) or search engine (VLOSE) – so designated under the DSA[58]– then they are subject to the risk-management framework provided for in the DSA, and there is a presumption that the obligations of the AI Act are fulfilled.[59] Going further down the rabbit hole, one assumes that if a generative AI system is embedded in an "online content-sharing service provider" ,[60] then not only would the DSA apply but also the special regime of Article 17 CDSMD, at least as regards certain copyright content moderation obligations on the output side.[61]

Why might these distinctions matter? On the one hand, the distinction between model provider and upstream players like Common Crawl and LAION (when acting merely as a dataset and tools provider) are consequential. Simply put, when these entities carry out TDM, they are subject to CDSMD requirements but not to the AI Act's copyright-relevant obligations.

What about the distinction between model and system? To be sure, if we consider the copyright obligations in the AI Act to be only ex ante obligations that target the model provider when building the model, then this distinction might have little practical relevance. However, there are two scenarios that warrant caution against this conclusion. First, it is possible–although not certain–that the "policies to respect copyright" obligation applies also to ex post moderation of outputs. Second, the potential copyright harm that the AI Act tries in part to address might come from how a *GPAI system* is deployed, e.g. in publicly available and widely used image, music or text generators. In that case, the act of infringement occurs only at the stage of deployment of the AI system incorporating the model. From the perspective of copyright liability, the model is merely a means to generate outputs downstream. Therefore, the GPAI model provider would at most be subject to secondary liability claims at national level vis-à-vis the infringing output.[62] In these scenarios, the model vs system distinction matters.

On the second common feature, because the copyright obligations in the AI Act are separate and independent from those in the copyright *acquis*[63] – namely, the TDM exceptions in the CDSMD – it follows that the failure by a GPAI model provider to comply with such obligations does not necessarily lead to an assessment of copyright infringement. Rather, they will lead to (possibly hefty) administrative fines. Some *regulatory spill-over* between the AI Act and the CDSMD may occur, as further explained below.[64]

Third, the copyright obligations of the AI Act are of broad application, cutting through some of the exceptions in the Regulation. In particular, they also apply to GPAI models released under a free and open-source license, and they apply to all types and sizes and models.[65] Compliance with these obligations should be commensurate and proportionate to the type of model provider, and should allow simplified ways of compliance for SMEs, including start-ups. This means, for instance, that for cases of modification or fine-tuning of a model, the relevant obligations should be limited to that modification or fine-tuning, e.g. by complementing and updating existing documentation.[66]

Finally, these obligations are to be monitored by the AI Office, a part of the European Commission.[67] In other words, this is a matter of compliance to be monitored by a public body rather than an avenue for private enforcement by copyright holders, which is considered as a separate issue.

4.3. Policies to respect copyright law, especially TDM opt-out

In general, this obligation requires a *policy* to respect EU copyright law. It is not clear whether this means the same as an obligation of result to respect EU copyright law. Instead, it appears to be at most an obligation of means or even only a formal requirement to have a copyright policy in place. Time will tell which is the correct interpretation, but the last option would seem an unsatisfactory outcome and partly at odds with recital 106 AI Act, which mentions *compliance* with the reservation of rights in Article 4(3) CDSMD.

The general obligation here relates to compliance with EU copyright law. As such, theoretically at least, it is not limited to the input or training stage but could also conceivably apply to all stages of the AI model value chain, possibly even to downstream moderation of outputs.[68] However, if copyright infringing outputs are mostly generated once the GPAI model has been integrated into a system, it remains unclear how to operationalize this in the AI value chain beyond the training stage.

Moving forward, the main focus of this obligation is to *identify* and *respect*, through state-of-the-art technologies, the reservations of rights

[57] Article 53(1)(b) AI Act.

[58] European Commission, "Supervision of the Designated Very Large Online Platforms and Search Engines under DSA" (2024) <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses> accessed January 27, 2025.

[59] Recital 118 AI Act. NB Article 3(65) AI Act defines "systemic risk" only in connection to GPAI models. The DSA, for its part, does not set out a definition of "systemic risks" but contains additional obligations for providers of very large online platforms and of very large online search engines to manage systemic risks in Articles 33 to 43. Importantly, Article 34(1) DSA itemizes types of systemic risks. For an analysis of whether the systemic risk assessment mechanisms set out in the DSA and AI Act may be used to address unauthorised use of protected content in the AI training phase, see Abbamonte (n 2) 486–487. On the topic of the interface between AI Act and DSA, see also Paddy Leerssen, "Embedded GenAI on Social Media: Platform Law Meets AI Law - DSA Observatory" (*DSA Observatory Blog*, October 16, 2024) <https://dsa-observatory.eu/2024/10/16/1864/> accessed January 13, 2025.

[60] As defined in Article 2(6) CDSMD.

[61] On the relationship between the DSA and the CDSMD, see e.g. Peukert and others (n 45); Quintais and Schwemer (n 45); João Pedro Quintais and others, "Copyright Content Moderation in the European Union: State of the Art, Ways Forward and Policy Recommendations" [2024] IIC - International Review of Intellectual Property and Competition Law <https://doi.org/10.1007/s40319-023-01409-5> accessed January 27 , 2025.

[62] NB secondary liability for copyright infringement is not harmonized in EU law, which merely established liability exemptions or "safe harbors" for intermediary service providers , initially in articles 12 to 14 e-Commerce Directive and currently in articles 4 to 7 DSA. See Christina Angelopoulos, "Harmonising Intermediary Copyright Liability in the EU: A Summary" in Giancarlo Frosio (ed), *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press 2020) <https://papers.ssrn.com/abstract=3685863> accessed January 27, 2025; Folkert Wilman, "The Digital Services Act (DSA) - an Overview" (December 16, 2022) <https://papers.ssrn.com/abstract=4304586> accessed January 27, 2025; Folkert Wilman, "The EU's System of Knowledge-Based Liability for Hosting Service Providers in Respect of Illegal User Content – between the e-Commerce Directive and the Digital Services Act" (2021) 12 JIPITEC.

[63] Recitals 108, 109 AI Act.

[64] See infra at 4.6 and 4.7.

[65] See Recital 104 and Article 53(2), not applying to paras (c) and (d)).

[66] Recital 109 AI Act.

[67] Recital 108 AI Act.

[68] See discussion above at 4.3.

(opt-out) in Article 4(3) CDSMD. In other words, the main target of the obligation appears to be operationalizing the opt-out mechanism in the context of TDM for developing commercial generative AI models. The reference to state-of-the-art technologies appears to be a call for the use of technical standards to opt-out, something that is confirmed by Article 53(4) AI Act.[69]

A few additional remarks on the opt-out mechanism are warranted. Article 4(3) CDSMD applies on condition that the use of protected material "has not been expressly reserved by their rights holders in an appropriate manner, such as machine-readable means in the case of content made publicly available online".[70] Recital 18 of the directive provides interpretative guidance by clarifying that:

- For content "made publicly available online", the only appropriate means to reserve (opt-out) is via machine-readable means, "including metadata and terms and conditions of a website or a service";
- In other cases, other means may be appropriate, e.g. contractual agreements or unilateral declaration;
- Copyright holders "should be able to apply measures to ensure that their reservations in this regard are respected".

There are significant questions about how to apply this provision in the training stage of GPAI models: is the object of the opt-out the work or each of its digital copies; can opt-out occur at any moment during the training stage or only when content is publicly available online (e.g. during web scraping or harvesting)? Can the opt-out occur at source page-level and/or at the training data-level? Is the opt-out conceptually a part of the right of reproduction of the copyright holder or is it dependent on acts of TDM? Relatedly, what is the relevant legal geographic point of attachment to exercise an opt-out? Finally, what constitutes an adequate "machine readable" opt-out?

The latter question was initially at the center of the recent *Kneschke v LAION* proceedings, but was eventually preempted by the qualification of LAION as a "research organization", meaning that its TDM activities fall under the national equivalent to Article 3 CDSMD, which does not require opt-out.[71] Despite this, the Hamburg court does mention as *obiter dicta* that it would consider opt-outs via natural language on terms of service as sufficient to meet the legal requirements of Article 3 CDSMD.[72] In my view, such a conclusion would be contrary to Recital 18 CDSMD, which indicates at the very least that an opt-out via terms and conditions of a website is only appropriate if it is also machine readable, meaning that additional requirements of format, metadata, clarity, and rights information should be necessary in order to meet this threshold.[73] This is also consistent with the notion that the AI Act aims for creation of technical standards for opt-out, which will clearly go beyond the presentation of a rights reservation in natural language in the terms and conditions of a website.

In any case, a key aspect here is that the requirement of opt-out in the CDSMD was not conceived with GPAI model providers in mind. Rather, for content made publicly available online, it seems to be aimed primarily at website owners where protected content is hosted and at the entities that carry out scraping or harvesting of those websites. In some cases, these entities will be GPAI model providers. In other cases, like with Common Crawl or LAION, they will not. The relevant AI Act obligations, for their part, target only GPAI model providers. Therefore, whatever technical solutions are developed in the context of compliance with the AI Act, they will not apply to other entities doing TDM and subject to the opt-out requirement, unless by coincidence they also qualify as GPAI model providers.

Now, bearing this in mind, there are already some relevant practical examples in the field of how opt-out could be operationalized. Keller and Warso, for instance, provide a good list of "TDM Reservation protocols".[74] Current well-known examples of opt-out approaches include Spawning AI's suite of tools[75] for rights holders (Have I Been Trained,[76] Kudurru,[77] or even a Browser Extension[78]), Google-Extended,[79] OpenAI's announced Media Manager,[80] and Cloudflare's crawl blocker.[81] This development of third-party and own provider tools is similar to what we saw happen for online content-sharing platforms (think YouTube's ContentID vs Pex tools).[82] In my view, this opens the door for the possible future use of the same tools beyond the training stage and well into downstream output filtering as a means to enforce copyright.[83] It remains to be seen how to articulate such an approach with the labeling obligation imposed on certain providers of AI *systems* (including GPAI systems) generating synthetic audio, image, video or text content, to ensure that the outputs of such system "are marked in a machine-readable format and detectable as artificially generated or manipulated".[84]

[69] See infra at 4.6.

[70] See generally on the topic of opt-out, highlights questions about its timing, mode and scope, Péter Mezei, "A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI" (2024) 46 European Intellectual Property Review <https://papers.ssrn.com/abstract=4695119> accessed 27 January, 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" (2024) 15 JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law <https://www.jipitec.eu/jipitec/article/view/407> accessed January 13, 2025.

[71] See above at 2 and sources cited therein.

[72] Pukas and Nordemann (n 18); Goldstein, Stuetzle and Bischoff (n 16).

[73] See Abbamonte (n 2). drawing a parallel with the transparency requirement for terms and conditions in Article 14(5) DSA. See also Andres Guadamuz, "LAION Wins Copyright Infringement Lawsuit in German Court" (*TechnoLlama*, September 28, 2024) <https://www.technollama.co.uk/laion-wins-copyright-infringement-lawsuit-in-german-court> accessed October 10, 2024. noting the "quite controversial" conclusion of the German court in this regard.

[74] Paul Keller and Zuzanna Warso, "Defining Best Practices for Opting out of ML Training" (Open Future Foundation 2023) <https://openfuture.eu/publication/defining-best-practices-for-opting-out-of-ml-training> accessed July 31, 2024. See also "TDM Reservation Protocol (TDMRep)" (W3C Community Group 2024) W3C Community Group Final Report <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/> accessed July 31, 2024.

[75] See https://spawning.ai/

[76] See https://spawning.ai/have-i-been-trained

[77] See https://kudurru.ai/

[78] See https://spawning.ai/browser-extension

[79] See, Google, Overview of Google crawlers and fetchers (user agents), https://developers.google.com/search/docs/crawling-indexing/overview-google-crawlers

[80] OpenAI, "Our Approach to Data and AI" (*OpenAI*, May 7, 2024) <https://openai.com/index/approach-to-data-and-ai/> accessed July 31, 2024. For criticism, see Bernd Justin Jütte, "Open AI's Vison for a Social Contract – of Things to Come…" (*Kluwer Copyright Blog*, June 3, 2024) <https://copyrightblog.kluweriplaw.com/2024/06/03/open-ais-vison-for-a-social-contract-of-things-to-come/> accessed July 31, 2024.

[81] Sam Rhea, "Start Auditing and Controlling the AI Models Accessing Your Content" (*The Cloudflare Blog*, September 23, 2024) <https://blog.cloudflare.com/cloudflare-ai-audit-control-ai-content-crawlers> accessed October 10, 2024.

[82] See, respectively, https://support.google.com/youtube/answer/2797370?hl=en (ContentID) and https://pex.com/ (Pex). For a critical analysis of the legal issues with this tools in the context of copyright content moderation, see Martin Senftleben, João Pedro Quintais and Arlette Meiring, "How the EU Outsources the Task of Human Rights Protection to Platforms and Users: The Case of UGC Monetization" (2024) 38 Berkeley Technology Law Journal <https://papers.ssrn.com/abstract=4421150> accessed January 27, 2025.

[83] As suggested e.g. by Arvind Narayanan and Sayash Kapoor, "Generative AI's End-Run around Copyright Won't Be Resolved by the Courts" (*AI Snake Oil*, March 20, 2023) <https://www.aisnakeoil.com/p/generative-ais-end-run-around-copyright> accessed January 27, 2025.

[84] Article 50(2) AI Act.

Setting aside for now the potential freedom of expression concerns arising from that future scenario, the main difference here is that the AI Act clearly pushes for standardization of opt-outs.

In that sense, it is not clear that existing approaches will be in line with what will be ultimately required by the AI Act, which in my view trends towards a more horizontal approach, possibly linked to some form of public registry infrastructure.[85]

What seems clear is that most of the public "opt-out" declarations by a number of collecting societies and rights holders all over Europe are not "appropriate" in the legal sense, even if they might be useful for political signaling purposes.[86]

4.4. Policies, opt-out and extraterritorial effect

One unexpected component of this "policies to respect copyright" obligation is found in Recital 106. According to this, the obligation should apply even if the TDM activities in question take place outside the EU, for instance in a jurisdiction with laxer requirements. The rationale is that such a rule is "necessary to ensure a level playing field" among GPAI model providers "where no provider should be able to gain a competitive advantage in the Union market by applying lower copyright standards than those provided in the Union." In other words, the aim is to prevent regulatory arbitrage. This follows a product safety logic that is consistent with the spirit of the AI Act: if you place a product (model) on the market in the EU, it should comply with EU law.

Abbamonte labels this as a "market entry requirement" that is derived from the general rule in Article 2(1)(a) AI Act, according to which the Regulation applies inter alia to providers placing on the market GPAI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country.[87] This is essentially also the argument advanced by Stiepe and Denger, which consider it sufficient to conclude that as a result EU copyright directives "acquire direct effect vis-à-vis private legal entities outside the EU".[88] Color me skeptical.

Assuming the "copyright policies" obligation means more than a formal requirement to put a policy in place,[89] there are at least two problems with this extraterritorial provision. These are not related to the normative desirability of the rule, but rather to its descriptive legal interpretation. First, the provision is contained in a recital. Recitals are not binding and their primary function in EU law is "to explain the essential objective pursued by the legislative act".[90] As the Court has consistently stated, recitals cannot directly create rights or duties.[91] Therefore, recitals cannot be of a norm-setting character.[92] In my view, recital 106 does not aim to clarify the application of the general market location approach of the AI Act to substantive copyright rules. Rather, the recital is intended to support the interpretation of the provision containing the obligation to which it explicitly refers, located in Article 53(1)(c). In doing so, it establishes an additional norm on extraterritoriality that extends beyond the enacting term it references. In other words, the normative exhortation in recital 106 exceeds the legal provision it supports, as it broadens the territorial scope of the copyright policy obligation—and by extension, substantive copyright rules—particularly the opt-out obligation.

This leads us to the second and related problem: the territoriality principle of copyright law and the application of the rule of *lex loci protectionis* (the law of the country for which protection is claimed)[93] to TDM activities. The territoriality principle binds the grant and effect of copyright to the territory of the state where protection is conferred to a work. Protection is granted on a territory-by-territory basis, with the attached exclusive rights following the same logic. It is therefore crucial to localize the relevant restricted act, for instance the reproduction of a work for TDM purposes. In the EU, Article 8 of the Rome II Regulation deals with conflict of laws regarding non-contractual obligations.[94] This provision clarifies that the *lex loci protectionis* applies to infringement of copyright, covering both the requirements and scope of protection.[95]

This means, in short, that if I carry out the relevant TDM acts (reproductions and extractions) to pre-train and train the GPAI model outside the EU, then the law applicable to those acts is the law of the place where those reproductions and extractions take place, *not the law where the trained GPAI model is subsequently made available*. If such a place is not in the EU, then the national laws of Member States implementing the TDM exceptions are not applicable. As a result, there is no infringement of Article 4 CDSMD if the model is only placed on the EU market *post-training*, at a stage where no further TDM takes place[96]

Peukert calls this a "minimalist" solution, as opposed to the "maximalist" approach of postulating the extraterritorial application of the AI Act in disregard of the principle of copyright territoriality.[97] He then advances an "intermediate solution", which consists of "making the application of Art. 53(1)(c) AI Act dependent on whether the model provider scraped websites hosted on servers located in the EU."[98]

Arguably, however, this would already result from the application of

[85] As e.g. suggested by Paul Keller, "Considerations for Implementing Rightholder Opt-Outs by AI Model Developers" (Open Future Foundation 2024) <https://openfuture.eu/publication/considerations-for-implementing-rightholder-opt-outs-by-ai-model-developers> accessed July 31, 2024. See also Abbamonte (n 2) 483.

[86] See, e.g., https://societe.sacem.fr/actualites/notre-societe/pour-une-intelligence-artificielle-vertueuse-transparente-et-equitable-la-sacem-exerce-son-droit (SACEM), https://www.sabam.be/en/press/sabam-safeguards-rights-its-authors-ai-use (Sabam), https://pictoright.nl/nieuws/collectieve-opt-out-pictoright-aangeslotenen/ (Pictoright), https://bumastemra.nl/bumastemras-uitgebreide-standpunt-over-ai/ (BumaStemra), and https://www.sonymusic.com/sonymusic/declaration-of-ai-training-opt-out/ (Sony).

[87] Abbamonte (n 2) 285.

[88] Malte Stieper and Michael Denga, *The international reach of EU copyright through the AI Act* (Institut für Wirtschaftsrecht 2024) 14 <https://opendata.uni-halle.de//handle/1981185920/118909> accessed October 22, 2024.

[89] See discussion above at 4.3

[90] Maarten den Heijer, Teun van Os can den Abeelen and Antanina Maslyka, "On the Use and Misuse of Recitals in European Union Law" (SSRN, 2019) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3445372> accessed July 31, 2024. See also Legal service (European Commission), *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) 31–36 <https://data.europa.eu/doi/10.2880/5575> accessed October 30, 2024; Todas Klimas and Jurate Vaiciukaite, "The Law of Recitals In European Community Legislation" (2008) 15 ILSA Journal of International & Comparative Law 61.

[91] Case C-136/04, Deutsches Milch-Kontor v Hauptzollamt Hamburg-Jonas, EU:C:2005:716, para 23; Case C-134/08, Hauptzollamt Bremen v J. E. Tyson Parketthandel, EU:C:2009:229, para 19.

[92] den Heijer, can den Abeelen and Maslyka (n 90) 6; Legal service (European Commission) (n 90).

[93] Recently on the topic, see Mireille MM van Eechoud, "Territoriality and the Quest for a Unitary Copyright Title" (2024) 55 IIC - International Review of Intellectual Property and Competition Law 66.

[94] 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.

[95] Stieper and Denga (n 88).

[96] See, similarly, Peukert (n 17).

[97] ibid 9–12.

[98] ibid 11–12. (identifying problems with the "lex scraping" approach).

the normal minimalist approach in accordance with the principles of territoriality and *lex loci protectionis*. In my view, if any of the TDM activities has a clear point of attachment with EU territory–most notably web scraping–then the model provider should have to respect EU copyright law, including the opt-out requirement.[99]

If the entity carrying out the scraping is the GPAI model provider, then it will have to comply not only with the requirements of Article 4(3) CDSMD but also with the additional requirements of Article 53(1)(c) AI Act. However, it may well occur that the entity that carried out the relevant TDM activity is not a GPAI model provider, as in the case of Common Crawl (for web scraping) and LAION (for dataset preparation). In such capacity, these entities are not subject to the obligations in the AI Act. As such, it is difficult to envisage how GPAI model providers can ensure an effective opt-out for content and datasets lawfully scraped or prepared by upstream third parties.

Another defense of the AI Act's extraterritorial effect on copyright issues is advanced by Rosati. Building on examples from international and EU copyright law, including CJEU case law on localization of copyright infringing acts, she argues that if the acts of extraction and reproduction during TDM are "functionally" essential to the training of AI models, and those models are made available for use in the EU, it is justified to apply EU law to these acts, considering they are part of a broader process connected to the EU.[100]

But this interpretation is difficult to reconcile with the practice and rules applicable to TDM. There is a clear factual and legal distinction between (1) TDM activities required to train and build a model, and (2) the making available of that model on the EU market.[101] These activities are treated differently under EU copyright law, where it is not possible to conflate the legal regime of TDM – which applies to rights of reproduction and extraction – and the act of making available of a trained GPAI model.[102] Furthermore, these activities are also distinct in the logic of the AI Act, which demarcates the training of a GPAI model from the subsequent integration of the trained model in an "AI system", its "placing on the market", "making available on the market", and its "putting into service".[103] In my view, no amount of functional interpretation can cross this particular interpretive Rubicon.

In sum, if the TDM leading up to the model took place outside the EU, then *EU copyright law* does not require GPAI model providers to ensure that the resulting model complies with Article 4 CDSMD. Therefore, even if this recital is turned into a binding obligation by national law, its violation does not amount to copyright infringement. It would only be a violation of the AI Act, subject to the sanctions and penalties mentioned below. Even then, since this particular obligation refers back to the "policies to respect copyright" obligation, it seems odd to impose a sanction on a provider for failing to comply with EU copyright law when that provider has, in fact, respected the applicable copyright rules.[104] It seems even stranger to recognize such a deviation from the core principles of EU copyright law based on a recital in a legislative instrument that only tangentially relates to copyright.

One possible solution to this issue is explored in the draft Code of Practice for GPAI model providers. The compliance mechanism of the Code of Practice is detailed below in section 4.6. Here, it is worth noting the following key points. The logic of the Code is to first establish general commitments, followed by concrete measures and specific key performance indicators (KPIs) to monitor and ensure the effectiveness of these measures.

Rather than imposing mandated extraterritoriality, the Code suggests a form of voluntary extraterritoriality. Under this approach, GPAI model providers would agree to make a model available on the EU market only if the model complies with EU copyright law throughout its value chain or lifecycle. For instance, Measure 2.3 seeks to reinforce this logic by emphasizing due diligence and contractual operations, addressing the non-applicability of the Act's copyright obligations to upstream providers. This measure stipulates that providers should make reasonable efforts to assess the copyright compliance of third-party datasets used in the development of GPAI models, including verifying whether these third parties respected TDM opt-outs under Article 4(3) CDSMD. Notably, this commitment would extend to cases where such TDM activities by third parties occurred outside the EU.[105]

Furthermore, the Code avoids restricting the application of copyright-relevant measures to TDM activities conducted solely within the EU. This intent is evident in Measure 2.5, where signatories commit to taking reasonable steps to exclude pirated sources from their crawling activities, including those listed in piracy watchlists published by relevant public authorities in jurisdictions where GPAI model providers are established, including those outside the EU.[106]

In my view, this approach to voluntary extraterritoriality, achieved through meta-regulation with robust policy and market incentives, aligns with EU copyright law. It also offers a more promising avenue for ensuring compliance with the AI Act's "copyright policies" obligation.

4.5. Transparency

The second key copyright obligation in the AI Act is in fact broader than copyright. Article 53(1)(d) states that GPAI model providers must draw up and make publicly available a sufficiently detailed summary about the data that is used in the training (including pre-training) of GPAI models, according to a template provided by the AI Office. Importantly, this obligation covers data beyond copyright-protected content and legitimate interests beyond those of copyright holders, such as those related to privacy and data protection, research, the prohibition of discrimination and respect for diversity (related e.g. to the avoidance or mitigation of bias), and fair competition.[107] That is to say, copyright interests should not be the only ones considered in shaping the contours of the obligation.[108]

[99] For different interpretative options see Peukert (n 2); Senftleben (n 32).
[100] Rosati, "Infringing AI" (n 26).
[101] See e.g. Stieper and Denga (n 88). treating training, design and output as "separate regulatory subjects".
[102] ibid. ("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.")
[103] Cf. inter alia the definitions in Article 3 (9)-(11), (63), (66).
[104] Contra see Stieper and Denga (n 88). (reaching a different conclusion and arguing for a broad extraterritorial effect of EU copyright law by means of the AI Act based on justifications related to securing a "European level playing field", the "reversal of diffusion through intermediary liability", the fact that providers are the cheapest cost avoider, and the logic of product safety law).
[105] Second Draft General-Purpose AI Code of Practice, p.22, available at https://digital-strategy.ec.europa.eu/en/library/second-draft-general-purpose-ai-code-practice-published-written-independent-experts (accessed 13 January 2025).
[106] Second Draft General-Purpose AI Code of Practice, pp.22–23 , available at https://digital-strategy.ec.europa.eu/en/library/second-draft-general-purpose-ai-code-practice-published-written-independent-experts (accessed 13 January 2025).
[107] Zuzanna Warso, Maximilian Gahntz and Paul Keller, "Sufficiently Detailed? Towards Robust Training Data Transparency" (Open Future Foundation; Mozilla 2024) <https://openfuture.eu/publication/towards-robust-training-data-transparency> accessed January 27, 2025. On the topic of bias and copyright, see also Thomas Margoni, João Pedro Quintais and Sebastian Felix Schwemer, "Algorithmic Propagation: How the Data-Platform Regulatory Framework May Increase Bias in Content Moderation, Forthcoming in: Caterina Sganga & Tatiana Eleni Synodinou (Eds), , 2025
" in Caterina Sganga and Tatiana Synodinou (eds), *Flexibilities in Copyright Law, Routledge* (Routledge 2025) <https://papers.ssrn.com/abstract=4913758> accessed January 27, 2025.
[108] Warso, Gahntz and Keller (n 107); Zuzanna Warso, Maximilian Gahntz and Paul Keller, "Blueprint of the Template for the Summary of Content Used to Train General-Purpose AI Models (Article 53(1)d AIA) – v.2.0" (Open Future Foundation; Mozilla 2024) Policy Brief <https://openfuture.eu/blog/sufficiently-detailed-summary-v2-0-of-the-blueprint-for-gpai-training-data/>.

As to the contents of this obligation, it appears clear that the goal cannot be for GPAI providers to list all or most of the copyrighted material they are including in their training data sets in an itemized manner with clear identification of rights ownership claims, etc., as this would make this provision impossible to comply with. The low threshold of originality, the territorial fragmentation of copyright and its ownership, the absence of a registration requirement for works, and in general the poor state of rights ownership metadata[109] demonstrate this impossibility.

Rather, as the AI Act clarifies, the summary in question must be generally comprehensive in its scope instead of technically detailed to facilitate exercise and enforcement of rights, e.g. by copyright holders. The summary must include both copyrighted and non-copyrighted content, and it must be drafted taking into consideration the protection of trade secrets and confidentiality. Examples of the content of such a summary include listing main data collections or sets for training the model (e.g. large private or public databases or data archives) and providing a narrative explanation about other data sources used.[110]

The AI Office is tasked with providing the template that meets these requirements, as well as to monitor compliance with these obligations. Importantly, there should be no need for the Office to carry out a work-by-work assessment of the training data in terms of copyright compliance.[111]

From a copyright perspective, this obligation may help clarify three aspects that are essential to determine compliance with the TDM exception in Article 4 CDSMD. First, it may help establish whether relevant TDM activities took place in – or have sufficient points of attachment with –the EU territory, in order to trigger the application of the relevant exclusive rights of reproduction and extraction and, in a subsequent step, the TDM exception. Second, assuming EU law applies, it will help determine whether the GPAI model provider complied with the lawful access requirement – namely in web scraping – in Article 4 CDSMD.[112] Third, by providing access to the training sources and possibly the period on which these were accessed and used, it may help determine whether the TDM activities at issue respected the opt-outs in place at that source and time.[113]

Finally, as to the contents of such a summary, Warso, Gahntz & Keller advance a promising blueprint for a template.[114] That blueprint, for instance, identifies as key categories to be considered those of overall size of the training data, fine-grained detail on data sets and data sources (including a breakdown of the origin of sources that is crucial to access their compliance with copyright rules) ,[115] data diversity, and data processing.[116] The upcoming Code of Practice should further specify these categories in a meaningful manner.

4.6. Compliance with key obligations: harmonized standards, codes of practice, and commission approval

Article 53(4) AI Act provides guidance on how GPAI model providers may comply with the two key copyright-related obligations above. In simple terms, the provision advances a three-stage compliance approach.

The gold-standard and endgame is to develop a European harmonized standard.[117] Compliance with this standard – e.g. for the transparency obligation or opt-out – would grant providers the presumption of conformity with the respective obligations.

Until such a standard is published, GPAI model providers may rely on codes of practice to demonstrate compliance with the obligations. Article 56 AI Act regulates such codes of practice, which are to be drawn-up with the AI Office acting as a facilitator. Article 56(2) then sets out an obligation for the AI Office and the Board to ensure that the codes of practice cover at a minimum certain obligations provided for inter alia in Article 53. Here, it is noteworthy that such minimum obligations only cover "the adequate level of detail for the summary about the content used for training".

At the time of writing of this paper, discussions on a GPAI code of practice are underway.[118] The Code of Practice will detail the AI Act rules for GPAI model providers, including those with systemic risks. These rules will apply twelve months after the entry into force of the AI Act. The code of practice discussion takes place in the code of practice Plenary, which is structured in four working groups on specific topics. The Plenary kick-off meeting took place on 30 September 2024. The working groups, including a specific one on transparency and copyright-related rules, will meet several times for drafting rounds between September 2024 and April 2025. The final version of the first code of practice will be presented in a closing Plenary, expected to take place in April 2025, and subsequently published.[119] It is noteworthy that although codes of practice are in principle "soft law", the Commission may, by way of an implementing act, approve a code of practice and give

[109] See e.g. Martin Senftleben and others, "Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive" (2022) 13 JIPITEC <https://www.jipitec.eu/issues/jipitec-13-1-2022/5515>; Geiger and Iaia (n 32).

[110] Recital 107 AI Act

[111] Recital 108 AI Act. "Training data" is defined in Article 3(29) as data used for training an AI system through fitting its learnable parameters.

[112] On which topic, see Thomas Margoni, "Saving Research: Lawful Access to Unlawful Sources under Art. 3 CDSM Directive?" (*Kluwer Copyright Blog*, December 22, 2023) <https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/> accessed January 27, 2025.

[113] Abbamonte (n 2) 484.

[114] Warso, Gahntz and Keller (n 107). (Disclosure: I was one of the academics providing feedback on this proposal).

[115] E.g. whether (and to what extent) the data was scraped from the internet, collected from public repositories, licensed from rightholders or third-party intermediaries, obtained from proprietary databases, generated by users of products or services offered by the provider, synthetically generated by the provider, generated by the provider by other means, or obtained through other means.

[116] See Warso, Gahntz and Keller (n 108). For a separate proposal built on case studies of existing AI models, see Thomas Heldrup, "Report on AI Model Providers' Training Data Transparency and Enforcement of Copyrights" (Danish Rights Alliance 2024) <https://rettighedsalliancen.com/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf>.

[117] See generally Josep Soler Garrido and others, "Harmonised Standards for the European AI Act" (European Commission 2024) JRC Report JRC139430 <https://publications.jrc.ec.europa.eu/repository/handle/JRC139430> accessed January 27, 2025.

[118] European Commission, "The Kick-off Plenary for the General-Purpose AI Code of Practice Took Place Online" (*European Commission - Shaping Europe's digital future*, September 30, 2024) <https://digital-strategy.ec.europa.eu/en/news/kick-plenary-general-purpose-ai-code-practice-took-place-online> accessed January 27, 2025. NB the author is one of the participants in the plenary discussions, in the role of independent expert.

[119] The Chairs for each of the four WGs of the Plenary, including WG1 on transparency and copyright-related rules are listed at European Commission, "Meet the Chairs Leading the Development of the First General-Purpose AI Code of Practice" (*European Commission - Shaping Europe's digital future*, September 30, 2024) <https://digital-strategy.ec.europa.eu/en/news/meet-chairs-leading-development-first-general-purpose-ai-code-practice> accessed 27 January, 2025.

it general validity within the Union.[120] Relatedly, the AI Act establishes that codes of practice must be ready at the latest by 2 May 2025; if that is not the case, the Commission may impose, through implementing acts, common rules for inter alia the copyright-related obligations in Articles 53.[121]

This leads us to the third compliance option in Article 53(3), where in the absence of an EU harmonized standard and if a GPAI model provider does not adhere to an approved code of practice (or none exists that covers a specific obligation), it must "demonstrate alternative adequate means of compliance for approval by the Commission". For opt-outs, this would mean that companies like OpenAI would have to go to the Commission and argue that a combination of their policies and technical solutions like OpenAI's Media Manager for creators and content owners[122] meet the key "copyright policy" obligation.

In my view, one important consequence of this regime is what I call "regulatory spill-over" between the (public law) AI Act and the (private law) EU copyright *acquis*, especially the CDSMD. Although these are separate instruments, the violation of which leads to different consequences, the AI Act obligations are clearly complementary to the assessment of core requirements of Article 4 CDSMD: lawful access (transparency requirement) and opt-out (policies requirement). As such, if a GPAI model provider is deemed to comply with these obligations in the AI Act, they will probably be presumed to comply with the lawful access and opt-out requirements in the CDSMD as well. But the contrary is not necessarily true, at least for the transparency obligation. The obvious illustration is where a GPAI model provider fails to provide a sufficiently detailed summary of its training data with there being no evidence that the model is trained on unlawfully accessed content. The opt-out obligation is trickier. In any case, since questions on the interpretation of provisions in the AI Act and the CDSMD can be subject to preliminary rulings from the CJEU it is likely that the Court will ultimately judicially harmonize much of the open questions surrounding the key copyright obligations of the AI Act, for instance by defining key terms therein as autonomous concepts of EU law.[123] In doing so, the Court should also clarify the nature of the relationship between the AI Act and Article 4 CDSMD.

4.7. Enforcement and remedies: public vs private

Article 88 AI Act gives the Commission exclusive powers to supervise and enforce compliance with the key copyright obligations of GPAI model providers in Article 53. Furthermore, the Commission shall entrust the implementation of these tasks to the AI Office.[124] In other words, enforcement of these obligations rests squarely on public bodies, namely the Commission and the AI Office, which have sprawling powers in this area. These powers include co-regulatory measures that may be of significant practical relevance, including the aforementioned codes of practice and standards. They also include, as a remedy set out in Article 93, the power by the Commission to request providers to take appropriate measures to comply with the copyright-relevant obligations set out in Article 53 and restrict the making available on the market, withdraw or recall the model.

What are the applicable sanctions or penalties? For negligent or intentional infringement of the copyright relevant obligations of the AI Act, Article 101 clarifies that the Commission may impose on GPAI model providers fines not exceeding 3 % of their annual total worldwide turnover in the preceding financial year or EUR 15 000 000, whichever is higher. An important aspect here is that in fixing the amount of the fine or periodic penalty payment the Commission must inter alia take into account commitments made by the GPAI provider in relevant codes of practice in accordance with Article 56. It is also noteworthy that the AI Act does not offer possibilities for private enforcement when its obligations are infringed. It only affords private parties the right to lodge a complaint with a market surveillance authority; in this case, the AI Office vis-à-vis GPAI model providers' obligations.

Although the potential fines are substantial, none of this money will line the pockets of copyright holders. In fact, for copyright holders, the main benefit of these AI Act provisions may be twofold. First, it might place pressure on GPAI model providers to comply with EU copyright law and thus, presumably, better design and deploy models to be copyright compliant and enter into licensing deals or other arrangements. This may already be happening, judging from the tsunami of agreements between providers, rights holders and aggregators. Examples include the controversial Microsoft deal with Taylor & Francis[125] and the multiple deals OpenAI has in place with players such as TIME,[126] the Atlantic, ,[127] the Financial Times,[128] Le Monde and Prisa Media,[129] Axel Springer,[130] and The Associated Press.[131132]

[120] Article 56(6) AI Act. On the different types of soft law (codes of conduct, condes of practice and standards) in the AI Act and their characterization as "meta-regulation", see 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, December 6, 2024) <https://papers.ssrn.com/abstract=4997886> accessed January 27, 2025. The authors qualify codes of practice as "embedded meta-regulation" and note that such codes in the AI Act "are probably compulsory…, at least as a 'stop-gap' measure prior to the adoption of harmonised standards". ibid 12, 20–21.

[121] Article 56(9) AI Act.

[122] OpenAI, "Our Approach to Data and AI" (n 80).

[123] On the CJEU's use of autonomous concepts of EU law and a general interpretative standard of EU copyright law provisions with a view to their harmonization, see Eleonora Rosati, *Copyright and the Court of Justice of the European Union | Oxford Academic* (2nd Edition, Oxford University Press 2023) 62–63 <https://academic.oup.com/book/57710>.

[124] European Commission, 'European AI Office' (*European Commission*, 2024) <https://digital-strategy.ec.europa.eu/en/policies/ai-office> accessed January 27, 2025.

[125] Wellett Potter, "An Academic Publisher Has Struck an AI Data Deal with Microsoft – without Their Authors' Knowledge" [2024] *The Conversation* <https://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203> accessed January 27, 2025.

[126] OpenAI, "Strategic Content Partnership with TIME" (*OpenAI*, June 27, 2024) <https://openai.com/index/strategic-content-partnership-with-time/> accessed https://openai.com/index/strategic-content-partnership-with-time/.

[127] OpenAI, "A Content and Product Partnership with The Atlantic" (*OpenAI*, May 29, 2024) <https://openai.com/index/enhancing-news-in-chatgpt-with-the-atlantic/> accessed January 27, 2025.

[128] OpenAI, "We're Bringing the Financial Times' World-Class Journalism to ChatGPT" (*OpenAI*, April 29, 2024) <https://openai.com/index/content-partnership-with-financial-times/> accessed January 27, 2025.

[129] OpenAI, "Global News Partnerships: Le Monde and Prisa Media" (*OpenAI*, March 13, 2024) <https://openai.com/index/global-news-partnerships-le-monde-and-prisa-media/?utm_campaign=The%20Batch&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-8LNS8DF2FAurlzNv-TFTZKbJ0jKgiLC0wmKt8MgCKBZQFvWmuJJuwXqSWNb-qAt3KNkO8m> accessed January 27, 2025.

[130] Axel Springer, "Axel Springer and OpenAI Partner to Deepen Beneficial Use of AI in Journalism" (*Axel Springer*, December 13, 2023) <https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism> accessed January 27, 2025.

[131] Associated Press, "AP, Open AI Agree to Share Select News Content and Technology in New Collaboration" (*The Associated Press*, July 13, 2023) <https://www.ap.org/media-center/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration/> accessed January 27, 2025.

[132] For a critical comment on some of these deals from the perspective of authors' interests, see Dave Hansen, "What Happens When Your Publisher Licenses Your Work for AI Training?" (*Authors Alliance*, July 30, 2024) <https://authorsalliance.substack.com/p/what-happens-when-your-publisher> accessed January 27, 2025.

Second, from the perspective of enforcement avenues, the "regulatory spillovers" mentioned above may help in establishing direct copyright infringement by a GPAI model provider, especially for failure to meet the requirements of the TDM exception in Article 4 CDSMD. If that is the case, then under EU law copyright holders will benefit from the array of measures made available in the EU Intellectual Property Enforcement Directive.[133] Contrary to the US, where statutory damages may present the greatest risk for model providers,[134] under EU law the possibility of injunctions may be the most threatening. For instance, the possibility of provisional or precautionary measures, corrective measures or even final injunctions aimed at prohibiting the continuation of the infringement might conceivably lead to unavailability or destruction of an infringing model on the EU market. Different flavors of model disgorgement, meaning injunctions for "the elimination or reduction of …any improperly used data [*and*] the effects of improperly used data on any component of an [*machine learning*] model" ,[135] could prove risky for GPAI model providers. At this stage, however, it is important to emphasize that model disgorgement methods and techniques (e.g. retaining, unlearning, compartmentalization) have yet to be properly developed and tested.[136] In judicial proceedings, their implementation will be challenging, not just from a technical perspective but also as regards the assessment of their proportionality vis-à-vis the type and level of infringement in the particular case.[137]

5. Remuneration: de lege lata and de lege ferenda

In all likelihood, the most consequential policy concern surrounding the debate on generative AI and copyright relates to how copyright holders – and specifically creators (authors and performers) – can and should be remunerated by the use of their works in the generative AI lifecycle or value chain. Overall, this has been a topic of much concern and debate in policy circles and academia, and was arguably the key motivator for the copyright related obligations in the AI Act. This is for instance a central point in the recent Joint Letter by European Composer & Songwriter Alliance (ECSA) to Members of the European Parliament on the impact of AI on the European creative community.[138] In academia, for its part, several proposals have been advanced to address this issue.[139] I first examine avenues for remuneration of creators in existing law (5.1) and then turn to representative proposals for legal reform (5.2). The purpose of this section is not to advance a standalone legal reform proposal but rather to provide interpretative options grounded in a critical analysis of the existing EU legal framework. In doing so, this section of the paper seeks to critically reflect on and contribute to debates on feasible approaches to fair remuneration of creators in the context of generative AI.

5.1. De lege lata

Looking at existing law, it can be said that the most obvious path to remuneration lies with leveraging the opt-out regime in Article 4 CDSMD, as supercharged by the obligations imposed by the AI Act on GPAI model providers. Importantly, these provisions target the input or training stage of AI model development. As a result, much of the debate at EU level focuses implicitly or explicitly on remuneration claims related to copyright relevant acts taking place during this stage. Still, it is undeniable that many of the concerns with the impact of generative AI on the creative sector are often expressed in connection with the potential that generative AI *outputs* have to compete with and negatively affect the commercial exploitation of creative works.[140] In other words, there is some conceptual dissonance between the scope of current EU law in this area and the commercial exploitation activities that copyright holders are most concerned with.

During the legislative process of the AI Act, there was some debate on whether relying on the opt-out model as a linchpin for creators' remuneration was a desirable development. On the one hand, some commentators considered that this approach had the potential to increase the bargaining power of rights holders and lead to licensing deals with (and remuneration from) AI providers.[141] In this vein, the civil society organisation Communia argued that the opt-out approach "constitutes a forward-looking framework for dealing with the issues raised by the mass scale use of copyrighted works for ML [*machine learning*] training […] it ensures a fair balance between the interests of rightholders on the one side and researchers and ML [*machine learning*] developers on the other".[142]

But there are also important criticisms of this approach. From an economic perspective, Martens views the copyright licensing of training data as problematic, arguing that it limits data quality and availability, increases costs, reduces competition, and slows innovation and productivity in generative AI, ultimately hampering economic growth.[143] From a contractual and remuneration perspective, Trendacosta and

[133] Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights (OJ L 157, 30.4.2004).

[134] For an overview of possible remedies in the US, see Pamela Samuelson, "How to Think About Remedies in the Generative AI Copyright Cases – Communications of the ACM" (*Communications of the ACM*, June 11, 2024) <https://cacm.acm.org/opinion/how-to-think-about-remedies-in-the-generative-ai-copyright-cases/> accessed January 27, 2025.

[135] See, e.g. Alessandro Achille and others, "AI Model Disgorgement: Methods and Choices" (2024) 121 Proceedings of the National Academy of Sciences e2307304121.

[136] See e.g. A Feder Cooper and others, "Machine Unlearning Doesn't Do What You Think: Lessons for Generative AI Policy, Research, and Practice" (December 1, 2024) <https://ui.adsabs.harvard.edu/abs/2024arXiv241206966F> accessed January 27, 2025.

[137] On the topic of proportionality and enforcement in this context, see e.g. Tuomas Mylly, "Regulating with Rights Proportionality? Copyright, Fundamental Rights and Internet in the Case Law of the Court of Justice of the European Union," *Copyright and Fundamental Rights in the Digital Age* (Edward Elgar Publishing 2020) <https://www.elgaronline.com/edcollchap/edcoll/9781788113878/9781788113878.00008.xml> accessed January 27, 2025; Lisa van Dongen, "Proportionality in IP Enforcement: A Tale of Two Frameworks" (2022) 2022 Intellectuele Eigendom en Reclamerecht 213.

[138] European Composer & Songwriter Alliance (ECSA), "Joint Letter to Members of the European Parliament on the Impact of Artificial Intelligence on the European Creative Community • News • ECSA - European Composer & Songwriter Alliance" (*ECSA - European Composer & Songwriter Alliance*, July 23, 2024) <https://composeralliance.org/news/2024/7/joint-letter-to-members-of-the-european-parliament-on-the-impact-of-artificial-intelligence-on-the-european-creative-community/> accessed January 27, 2025.

[139] See e.g. Senftleben (n 32); Geiger and Iaia (n 32).

[140] See, e.g. Senftleben (n 32); Abbamonte (n 2) 479. See also the recent Statement on AI Training Data, https://www.aitrainingstatement.org/ ("The unlicensed use of creative works for training generative AI is a major, unjust threat to the livelihoods of the people behind those works, and must not be permitted."). At time of writing, this statement has almost 20.000 signatories.

[141] Paul Keller, "Protecting Creatives or Impeding Progress? Machine Learning and the EU Copyright Framework" (*Kluwer Copyright Blog*, February 20, 2023) <https://copyrightblog.kluweriplaw.com/2023/02/20/protecting-creatives-or-impeding-progress-machine-learning-and-the-eu-copyright-framework/> accessed January 27, 2025.

[142] Communia, "Policy Paper #15 on Using Copyrighted Works for Teaching the Machine" (*COMMUNIA Association*, April 26, 2023) <https://communia-association.org/policy-paper/policy-paper-15-on-using-copyrighted-works-for-teaching-the-machine/> accessed January 27, 2025.

[143] Bertin Martens, "Economic Arguments in Favour of Reducing Copyright Protection for Generative AI Inputs and Outputs" (April 4, 2024) <https://www.bruegel.org/working-paper/economic-arguments-favour-reducing-copyright-protection-generative-ai-inputs-and> accessed January 27, 2025.

Doctorow argue that this approach would lead to market concentration and exploitation of creators by big companies.[144] Because creative labor markets are already heavily concentrated and dominant companies have significant bargaining power, they will be able to impose contractual terms on artists that require them to sign away their "training rights" for reduced compensation. The medium to long term result would be more concentration of power with large companies leaving less control and remuneration for artists.[145]

From a descriptive standpoint at least, some of this criticism is on point. It is already possible to observe on the market that commercial GPAI model providers are incentivized to enter into licensing deals with large rights aggregators to ensure high-quality datasets and avoid legal exposure.[146] So far, there is no evidence that these deals improve the remuneration of individual creators and artists.

This highlights a core issue not addressed in Article 4 CDSMD or the AI Act: creators can only opt-out of commercial TDM and subsequently enter into (individual or collective) licensing deals if they own the relevant reproduction rights. Otherwise, the parties to whom they transfer their rights in exploitation contracts will reap the benefit (if any) of licensing such "training rights."

De lege lata, this imbalance can only be partly redressed by considering training rights from the perspective of the rules on fair remuneration in exploitation contracts of authors and performers in Articles 18 to 23 CDSMD.[147] It is unclear whether these are a good fit in this scenario. In both the relationship with subsequent copyright holders/aggregators and with AI companies, there may be an important role for collective licensing and—more important still—for collective bargaining. However, this role is not without important barriers, even in a (so far hypothetical) world where the AI Act leads to standardization of opt-outs for TDM.[148]

From the perspective of collective licensing, several important obstacles emerge. These are worth highlighting, since they provide a measure of the challenge facing willing collective rights management organisations (CMOs) in this area.[149]

First, most CMOs operate on a territorial basis, meaning that they can only license uses taking place in their territory, or indirectly in other territories through sister CMOs with whom they have "representation agreements".[150] In other words, only if the TDM activities take place in the country of operation of the CMO are they entitled to provide a license.

Second, it is unclear whether CMOs have valid mandates for the "training rights" under discussion here. This might be because their agreements with members do not sufficiently include such a right, or because these rights have already been validly transferred on an exclusive basis by authors and performers to subsequent rights holders, like publishers and producers.

Third, the collective licensing approach might underestimate the monumental scale, complexity and cost of the rights clearance endeavor. Even in a world where the AI Act-powered transparency template clarifies with some granularity the sources used to train a GPAI model, it might not be feasible or cost effective for a CMO to license these works in any relevant EU territory. Consider that the now outdated GPT-3 was trained on approximately 570GB of filtered text data from a variety of sources (Common Crawl, WebText2, Books1, Books2, and Wikipedia), and that the dataset consisted of about 300 billion tokens.[151] This is an ocean of data where most CMOs' repertoire is but a drop. Unless the relevant TDM activities took place in the territory of the CMO and the organization represents a truly significant number of rights holders and works used to train a model, it is likely that the costs associated with the administration of the license (e.g. identification of works, collection and distribution) would be higher than the rights revenue to be distributed. For training data, if collective licensing is to be viable, a different approach will probably have to be devised, e.g. via packaging existing repertoires with high-quality metadata into datasets for model providers.[152]

Fourth and related, CMOs typically provide blanket licenses to their entire "repertoire".[153] However, to determine the scope of the repertoire they are allowed to license, CMOs will have to take into account the exercise of the opt-out mechanism in Article 4(3) CDSMD for each content item in that repertoire. It is unclear whether this is currently a feasible task from the operational perspective.

Fifth, EU law contains no compulsory mechanism of collective rights management for "training rights", meaning that only voluntary forms of collective management are possible.[154] The result is that CMOs can only license the repertoire from their members. The only exception to this would be if the CMO would be authorized to issue a collective license with extended effect (CLEE) for training rights, which would allow it to extend its representation powers to non-member rights holders, thereby offering a blanket licensing for training rights. The CDSMD sets forth a regime for measures to facilitate CLEE in Article 12, which include three possible licensing mechanisms: extended collective licensing; legal mandate; and legal presumption. According to this provision, CLEE can only: be managed by a CMOs; within well-defined areas of use; where direct individual licensing is "'typically onerous and impractical" ' (i.e. a market failure scenario); and in a way that "'safeguards the legitimate interests" of rightsholders. Article 12(3) further subjects the application of CLEE to a number of safeguards: sufficient representation, equal treatment, opt-out, and information obligations vis-à-vis rightsholders.

To the best of my knowledge, there is no CMO in the EU that has sufficiently met these requirements and is able to issue such licenses for training rights. In theory, that scenario would be possible in Member States and sectors where CLEE is well established, like the Nordic countries.[155] Despite that, it is difficult to imagine that this mechanism

[144] Katharine Trendacosta and Cory Doctorow, "AI Art Generators and the Online Image Market" (*Electronic Frontier Foundation*, April 3, 2023) <https://www.eff.org/deeplinks/2023/04/ai-art-generators-and-online-image-market> accessed January 27, 2025.

[145] ibid.

[146] See above at 4.7.

[147] Generally on the topic, see Quintais, "The New Copyright in the Digital Single Market Directive" (n 4); Giulia Priora, "Catching Sight of a Glimmer of Light: Fair Remuneration and the Emerging Distributive Rationale in the Reform of EU Copyright Law" (2020) 10 JIPITEC <https://www.jipitec.eu/issues/jipitec-10-3-2019/5043>; Dusollier (n 4).

[148] Raising legal and practical obstacles, see Senftleben (n 32) 1546–1549.

[149] Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market (CRM Directive). For an in depth analysis of this instrument, see Oleksandr Bulayenko and others, "Study on Emerging Issues on Collective Licensing Practices in the Digital Environment" (European Commission 2021) Study for the European Commission 3970490 <https://papers.ssrn.com/abstract=3970490> accessed January 27, 2025.

[150] See Article 3(j) CRM Directive.

[151] Tom B Brown and others, "Language Models Are Few-Shot Learners" [2020] arXiv:2005.14165 [cs] <http://arxiv.org/abs/2005.14165> accessed January 27, 2025.

[152] NB it is important to emphasize that the copyright-AI value chain goes beyond the training phase and also involves the possible use of protected works at the output stage, such as requests for news, which may also give rise to licensing deals. See e.g. some of deals struck by AI companies listed above at 4.7.

[153] See Article 3(l) CRM Directive.

[154] In this respect, the imposition of mandatory collective management or even a compensated exception for training rights (as a means to overcome a potential conflict with the norma exploitation) is conceptually difficult to reconcile with the possibility of rights holders to reserve their rights (opt-out) under Article 4 (3) CDSMD. See, similarly, Senftleben Senftleben (n 32) 1544–1546.

[155] At time of writing, there is a legislative proposal in Spain that attempts to apply CLEE in this area. See Teresa Nobre, "A First Look at the Spanish Proposal to Introduce ECL for AI Training" (*Kluwer Copyright Blog*, December 11, 2024) <https://copyrightblog.kluweriplaw.com/2024/12/11/a-first-look-at-the-spanish-proposal-to-introduce-ecl-for-ai-training/> accessed January 13, 2025.

would provide a EU-wide solution to collective licensing of training data.

Collective bargaining, for its part, is a largely underexplored in the copyright vs. AI discussion in Europe. However, it has been[156] (and continues to be[157]) used with some success in the US by authors' and performers' unions. On its face, it seems well-suited to deal with a range of issues that span the potential exploitation of copyrighted content in the AI value chain, from training to the generation of outputs and the use of AI tools in different creative sectors.

One possible point of entry for collective bargaining in EU copyright law is the regime of fair remuneration in exploitation contracts of authors and performers, in Articles 18 to 23 CDSMD.[158] Under this regime, Member States should be free to implement this principle of appropriate and proportionate remuneration in Article 18 CDSMD "through different existing or newly introduced mechanisms, which could include collective bargaining and other mechanisms, provided that such mechanisms are in conformity with applicable Union law."[159] Collective bargaining is also considered as an option for the relevant stakeholders to reach an agreement regarding the transparency obligation for authors and performers concerning the exploitation of their works and performances.[160] Finally, the existence of specific provisions in collective bargaining agreements plays a key role when assessing the contractual adjustment and rights of revocation mechanisms in Articles 20 and 22 CDSMD.[161]

More research is needed on how to leverage these rules–in tandem with collective licensing– to better remunerate creators for the use of their works in this context, as well as general rules on collective bargaining[162] to regulate the use of generative AI tools in the creative sectors that may be adversely affected by them.

5.2. *De lege ferenda*

De lege ferenda, new solutions would have to be introduced into EU law via legislative reform. These solutions could range from the qualification of training rights as unwaivable remuneration rights, to the introduction of AI levy systems or statutory licenses.[163]

These types of legal solutions echo previous calls by scholars and policy makers during the emergence of illegal file-sharing in the early 2000s for "alternative compensation systems". This is an umbrella term for legal schemes that authorise individuals to make online use of works or other protected subject matter – downloading, uploading, sharing, modifying – without the direct authorisation of rightsholders. This authorisation is subject to the payment of remuneration to creators or all rightsholders of protected works and subject matter included in the scheme.[164] The "alternative' refers to a deviation from the application of an exclusive right and individual management to large-scale online use of works. The previously restricted use becomes "permitted-but-paid" ,[165] signaling a shift from exclusivity to remunerated access and use.

Despite the name, most proposals for alternative compensation systems do not constitute radical departures from the status quo. Instead, they are extensions and adaptations of existing copyright regimes that may enable a model of access and remuneration in the online environment. Typically, inspiration is taken from collective rights management and legal licensing schemes.[166] The suggestions now emerging on how to deal with remuneration for creators in the context of generative can be seen as evolutions of these earlier proposals. They also provide creative solutions on how to address the dilemma of remuneration of creators while preserving AI innovation. For that reason, we examine below two representative examples of proposals that take as a departure point the inadequacy of the EU opt-out approach and focus on different phases of the generative AI lifecycle: Geiger and Iaia's statutory license for commercial TDM[167] and Senftleben's AI output levy system.[168]

Geiger and Iaia propose replacing the opt-out mechanism in Article 4 (3) CDSMD with a statutory license that includes a limitation-based remuneration right for commercial TDM activities outside the scope of Article 3 of the directive. This proposal draws from the private copying exception in the InfoSoc Directive and aims to balance the fundamental rights of rights holders and generative AI users.[169]

The proposal suggests two options for managing statutory licenses and remuneration for commercial generative AI TDM. First, it could involve a system of mandatory collective management, overseen by an EU-level or national supervisory authority to determine remuneration.[170] Second, legislators or appointed authorities would set criteria to fairly calculate compensation, based on the principles of appropriate and proportionate remuneration from Article 18 CDSMD, considering the economic value of the work and other factors, like relevant market practices in the sector.

Despite the positive aspects of this proposal, it faces significant obstacles. First, from the perspective of political feasibility, it is difficult to imagine the EU legislature being willing to implement the necessary measures to realize it. These measures include: (i) repealing the recently adopted opt-out mechanism in Article 4(3) of the CDSMD (and possibly amending relevant sections of the AI Act), (ii) passing new legislation to establish the new right and statutory licensing scheme, including the proposed mandatory collective management regime, and (iii) creating an EU-level supervisory body (or national-level bodies) for this purpose.

Second, in terms of legal feasibility, many aspects of the proposal would need clarification before a proper assessment can be made. These

[156] The Authors Guild, "SAG-AFTRA Agreement Establishes Important Safeguards for Actors Around AI Use" (*The Authors Guild*, January 17, 2024) <https://authorsguild.org/news/sag-aftra-agreement-establishes-important-ai-safeguards/> accessed January 27, 2025.

[157] Sarah Parvini, "Video Game Performers Will Go on Strike over Artificial Intelligence Concerns" (*AP News*, July 25, 2024) <https://apnews.com/article/sagaftra-video-game-performers-ai-strike-4f4c7d846040c24553dbc2604e5b6034> accessed January 27, 2025.

[158] NB CMOs are not collective bargaining organizations and are subject to specific rules in the CRM Directive. See e.g. recital 77 CDSMD.

[159] Recital 73 CDSMD.

[160] Article 19(5) and Recital 77 CDSMD.

[161] See Article 20(1) and 22(5), as well as recital 78 CDSMD.

[162] In EU law, Article 28 Charter sets out the right of collective bargaining and action. See also Communication from the commission Guidelines on the application of Union competition law to collective agreements regarding the working conditions of solo self-employed persons 2022/C 374/02, paras 37–40, clarifying inter alia that te CDSMD recognises "the right of certain solo self-employed to rely on collective agreements in order to correct an imbalance in bargaining power with their counterparty/-ies."

[163] See e.g. the proposals examined below by Senftleben (n 34); Geiger and Iaia (n 34). Other proposals include for instance Nicola Lucchi, "ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems" [2023] European Journal of Risk Regulation 1; Giancarlo Frosio, "Should We Ban Generative AI, Incentivise It or Make It a Medium for Inclusive Creativity?" in Enrico Bonadio and Caterina Sganga (eds), *A research agenda for EU copyright law* (Edward Elgar Publishing Ltd 2023); Daniel J Gervais, "The Remuneration Of Music Creators for the Use of Their Works by Generative AI" (International Council of Music Creators (CIAM); Fair Trade Music International (FTMI) 2024) White Paper <https://www.fairtrademusicinternational.org/wp-content/uploads/2024/08/FTMI-Gen-AI-White-Paper-2024.pdf>.

[164] See, for all, João Pedro Quintais, *Copyright in the Age of Online Access: Alternative Compensation Systems in EU Law* (Kluwer Law International 2017).

[165] Jane Ginsburg, "Fair Use for Free, or Permitted-but-Paid?" (2014) 29 Berkeley Tech. L. J. 1383.

[166] Quintais, *Copyright in the Age of Online Access: Alternative Compensation Systems in EU Law* (n 164).

[167] Geiger and Iaia (n 32).

[168] Senftleben (n 32).

[169] Geiger and Iaia (n 32).

[170] Ibid. The proposal for a EU level authority relies on Christophe Geiger and Natasha Mangal, "Regulating Creativity Online: Proposal for an EU Copyright Institution" (2022) 71 GRUR International 933. At a national level, inspiration is drawn from existing national bodies, like the AGCOM in Italy.

include, for instance, identifying the entities subject to the obligations imposed by the right: does it apply to the GPAI model provider, the AI system provider, other entities conducting TDM activities, or all of the above? Additionally, questions arise regarding territoriality, how to determine relevant TDM activities, and how to calculate and distribute remuneration across each Member State.

Third, all the legal and operational challenges related to collective licenses for training rights, highlighted above, apply here with even greater intensity, as the proposal assumes the creation of new legal constructs and rights management infrastructures. Some of these challenges are further elaborated below in relation to Senfleben's proposal, which envisions an even more innovative framework to ensure remuneration in the generative AI context.

Senftleben's proposal focuses on generative AI outputs and introduces a levy on providers of AI systems that produce content similar to literary and artistic works. Inspired by the residual remuneration right in Article 8(2) Rental Rights Directive, this regime would require AI system providers to pay remuneration for producing content with the potential to replace human creations. A general and abstract assessment of the replacement effect of AI outputs would trigger payment of a lump-sum levy, calculated as a percentage of providers' revenue. Legislators and judges would distinguish between AI systems that substitute human creativity and those that merely support it. The collected revenues would fund financial support, training, and new creative opportunities for authors, and would be distributed through CMOs.[171]

Crucially, the basis for the levy is not the use of copyright-protected content. Depending on a case-by-case assessment, generative AI outputs may or may not qualify as protected works under EU copyright law. These works may be authored by the prompter and, depending on the contractual relationship with the AI system provider, may be owned by the human prompter and/or the provider.[172] In some cases, the output may infringe the copyright of third parties whose works are included in the training data.[173] Some of these prima facie infringing uses may be permitted under exceptions in EU copyright law, such as parody, caricature, or pastiche. Consequently, from a copyright perspective, most outputs will not form the basis for a legal license justifying a levy, as they do not involve the mass use of third-party works without the authors' authorization. To address this issue, Senftleben draws on Dietz's concept of "domaine public payant", which advances "a new right to which a different right holder – the 'community of authors'– was entitled as a collective".[174] The key difference here is that the AI levy would provide copyright-like remuneration for "ideas, concepts, and styles" expressed as AI outputs, rather than for public domain works.[175]

Despite the merits of the proposal, I see at least four categories of obstacles. First, regardless of the analogy used, this is not a copyright levy. It lacks a sufficient link to the use of copyrighted works at the output level. Without that connection, the levy effectively becomes a tax, which is inconsistent with current EU copyright law.[176]

Second, given the lack of connection to copyright usage, the existing rules on TDM for training, and the applicability of copyright infringement rules to AI models *and* outputs, it is unclear what distinct harm the levy is meant to address. This raises both conceptual problems (how to align the levy with infringement rules?) and practical ones (how to fairly calculate a levy for harm that cannot be determined?). More importantly, it challenges the central normative justification for the AI levy: if the copyright harm cannot be identified, how can we establish the substitution effect of AI outputs on human-authored works?

Third, from a normative standpoint, this proposal would offer "para-copyright" protection—through an enforceable remuneration claim—to generally unprotected subject matter, such as machine-generated expressions of ideas, concepts, and styles (the AI output). This represents a significant departure from a fundamental principle of copyright law, the idea/expression dichotomy, while expanding the scope of copyright protection to non-original and non-human productions.[177]

Fourth, from the operational and implementation standpoints, this proposal faces challenges similar to those outlined above regarding collective licensing of training rights.[178] While there are important differences concerning the target of the levy (AI system provider vs GPAI model provider), the object (output vs input), the beneficiary (the "collective of authors"), and the rights management scheme (mandatory collective management, ideally through a currently non-existent pan-European CMO), most of these features would require significant changes to existing EU copyright law, as well as the development of new CMO infrastructure and practices. Combined, these obstacles lead to even greater political, legal and practical feasibility challenges than those observed for Geiger and Iaia's proposal.

These critiques are intended to be constructive, and these reform proposals should not be disregarded, as they offer thoughtful and creative solutions to the challenges generative AI poses to authors. First, they make a powerful case against the current opt-out approach. Second, they provide solid normative justification for a new approach that emphasizes fair author remuneration and respects the fundamental rights of various stakeholders. Third, they propose a blueprint for a solution based on the "collectivization" of rights exploitation by leveraging existing copyright mechanisms in the context of generative AI. Consequently, insofar as legal reform is feasible and aimed at protecting and fairly remunerating authors, lawmakers should give close consideration to these proposals, keeping in mind the substantive and operational shortcomings outlined above.

[171] Senftleben (n 32). NB the author examines the possibility of an input/training-based remuneration proposal but ultimately prefers an output-based approach.

[172] Hugenholtz and Quintais (n 25); João Pedro Quintais and Nick Diakopoulos, "A Primer and FAQ on Copyright Law and Generative AI for News Media" (*Medium*, April 26, 2023) <https://generative-ai-newsroom.com/a-primer-and-faq-on-copyright-law-and-generative-ai-for-news-media-f1349f514883> accessed January 27, 2025.

[173] Guadamuz (n 5); Rosati, "Infringing AI" (n 26).

[174] Senftleben (n 32) 1552. For the original proposal see Adolf Dietz, "A Modern Concept for the Right of the Community of Authors (Domaine Public Payant)" (1990) XXIV Copyright Bulletin 13.

[175] Senftleben further justifies this design on the fact that the systems were trained on copyright-protected content. Senftleben (n 32) 1554. However, in my view, this justification does not hold, since it conflates the different legal regimes and legal basis for remuneration applicable to training material and outputs.

[176] In the context of CJEU case law on the private copying levy and exception, the Court is clear that such a link needs to exist, despite the "rough justice" nature of the system. See Quintais, *Copyright in the Age of Online Access: Alternative Compensation Systems in EU Law* (n 164); Alexandre L Dias Pereira, "Levies in EU Copyright Law: An Overview of the CJEU's Judgments on the Fair Compensation of Private Copying and Reprography" (2017) 12 Journal of Intellectual Property Law & Practice 591.

[177] On generative AI, copyright and the protection of styles see e.g. Stephen Wolfson, "The Complex World of Style, Copyright, and Generative AI" (*Creative Commons*, March 23, 2023) <https://creativecommons.org/2023/03/23/the-complex-world-of-style-copyright-and-generative-ai/> accessed January 27, 2025; DJG Visser, "Stijlnabootsing met AI is onrechtmatig" (2023) 2023/2635 Nederlands Juristenblad <https://www.inview.nl/document/ide5f8f5c2f7e6499cb7ac00e1ad0d41df/nederlands-juristenblad-stijlnabootsing-met-ai-is-onrechtmatig> accessed January 27, 2025.

[178] See above at 5.1.

6. Conclusion

The emergence of generative AI has put pressure on EU copyright law. Although Article 4 CDSMD is likely the best tool currently available in the copyright *acquis* to address this issue, it falls short of being sufficient. The AI Act introduced two key obligations on GPAI model providers to supplement Article 4 CDSMD in the context of generative AI, focusing on transparency and the policies to respect EU copyright law. However, their effectiveness is debatable.

Beyond the overall complexity of the AI Act, its public law nature does not align well with the private law logic of copyright. This mismatch results in varied obligations, targets, a dubious extraterritorial effect, consequences for infringement, enforcement structures, and available remedies. The copyright-related obligations in the AI Act may positively influence the design of GPAI models with respect to copyright compliance and might even indirectly support the private enforcement of Article 4 CDSMD.

From the perspective of creators' remuneration, the AI Act further incentivizes GPAI model providers to enter into licensing deals with large rights aggregators to ensure access to high-quality datasets and avoid legal exposure. So far, however, these deals do not appear to improve the remuneration of individual creators and artists. A core issue with the existing regime is that creators can only opt out of commercial TDM and negotiate individual or collective licensing deals if they own the relevant "training rights." Otherwise, those rights will benefit larger rights holders and aggregators to whom they have transferred these rights in exploitation contracts. Existing rules in the CDSMD on fair remuneration (Articles 18–23) may only partially address this imbalance, and it remains unclear whether they are well-suited for the AI context. Collective licensing and collective bargaining could play an important role in resolving issues related to use of copyrighted content in the context of generative AI. However, collective licensing faces significant substantive and practical challenges that need to be addressed to make it a viable option in this area. Collective bargaining, while promising, remains relatively underexplored in EU copyright law.

Looking ahead, nuanced legal interpretation – along the lines suhested in this paper – and legislative reform are necessary to address these issues. Two reform proposals—Geiger and Iaia's statutory license for training and Senftleben's AI output levy—offer strong critiques of the current EU copyright rules for dealing with commercial generative AI via an opt-out system. They suggest reforms to improve creators' remuneration while balancing copyright protection, artistic freedom, and technological innovation. While these proposals face significant obstacles and are unlikely to be adopted in their current form, elements of them could be incorporated into future legal solutions to the challenges posed by generative AI for creators in the EU.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The author would like to thank Bernt Hugenholtz and Nuno Sousa e Silva for helpful comments on earlier drafts of this paper.

Data availability

No data was used for the research described in the article.