Generative AI, Copyright and the AI Act

João Pedro Quintais

Associate Professor
University of Amsterdam
Institute for Information Law (IViR)
j.p.quintais@uva.nl

v. 1.0

Contents

1. Introduction

This paper examines the copyright-relevant rules of the recently published Artificial Intelligence (AI) Act[1] for the EU copyright *acquis*.[2] The aim of the paper is to provide a critical overview of the relationship between the AI Act and EU copyright law, while highlighting potential gray areas and blind spots for legal interpretation and future policymaking.

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 text and data mining (TDM) exceptions in Articles 3 and 4 of the Copyright in the Digital Single Market Directive (CDSMD).[3] 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 offers some concluding remarks, focusing on the inadequacy of the current regime to address one of its main concerns: the fair remuneration of authors and performers.

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.[4]

---

[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*, 9 May 2023) <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/> accessed 18 August 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*, 11 December 2023) <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/> accessed 31 July 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 31 July 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] 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.

[4] 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.

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 mainly regulated by different exclusive rights of reproduction and the two TDM exceptions in the CDSMD.[5]

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." It is now mostly settled that this definition covers training activities needed to develop an AI model, including generative AI models. The AI Act reiterates this view.[6]

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, including lawful access.

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. 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.[7]

At this stage, most of the action in the EU is taking place at the legislative and policy level. But there is at least one relevant case making its way through German courts on the legality of using copyright-protected works for training generative AI models.[8]

The EU approach is different from that of other jurisdictions. 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

---

[5] 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.

[6] Se infra at 4.1

[7] See infra at 4.

[8] See Hamburg Regional Court, *Kneschke v. LAION*, discussed in Paul Keller, 'Machine Readable or Not? - Notes on the Hearing in LAION e.v. vs Kneschke' (*Kluwer Copyright Blog*, 22 July 2024) <https://copyrightblog.kluweriplaw.com/2024/07/22/machine-readable-or-not-notes-on-the-hearing-in-laion-e-v-vs-kneschke/> accessed 31 July 2024.

are more than 20 cases before US courts that tackle this question from different angles. [9] The current policy debate in Canada seems to toggle between the EU and US approaches.[10]

From the AI model perspective, the main legal questions that arise thus far relate to the whether the model weights qualify as protected databases[11] and whether there is sufficient memorization of protected content in a model for the same to qualify as containing (unauthorized) reproductions of that content.[12]

From the output perspective, key questions include whether AI-generated outputs are protected by copyright[13], whether outputs are derivative works, and if they infringe on third-party works used in training. 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.[14]

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 (see Articles 5(3)(d) and (k) InfoSoc Directive[15] and 17(7) CDSMD). In the EU, it will be particularly relevant to see whether and how the upcoming ruling on *Pelham II*[16] will qualify

---

[9] See 'Chat GPT Is Eating the World' (*Chat GPT Is Eating the World*) <https://chatgptiseatingtheworld.com/> accessed 31 July 2024. 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' <http://arxiv.org/abs/2309.08133> accessed 31 July 2024; Matthew Sag, 'Copyright Safety for Generative AI' (2023) 61 Houston Law Review <https://papers.ssrn.com/abstract=4438593> accessed 8 September 2023.

[10] Carys J Craig, 'Canada's Changing AI-Copyright Policy Discourse: A Play in Three Parts?' (*Kluwer Copyright Blog*, 25 April 2024) <https://copyrightblog.kluweriplaw.com/2024/04/25/canadas-changing-ai-copyright-policy-discourse-a-play-in-three-parts/> accessed 31 July 2024.

[11] Nuno Sousa e Silva, 'Are AI Models' Weights Protected Databases?' (*Kluwer Copyright Blog*, 18 January 2024) <https://copyrightblog.kluweriplaw.com/2024/01/18/are-ai-models-weights-protected-databases/> accessed 31 July 2024.

[12] A Feder Cooper and James Grimmelmann, 'The Files Are in the Computer: On Copyright, Memorization, and Generative AI' <http://arxiv.org/abs/2404.12590> accessed 31 July 2024; Ivo Emanuilov and Thomas Margoni, 'Forget Me Not: Memorisation in Generative Sequence Models Trained on Open Source Licensed Code' <https://zenodo.org/records/10635479/preview/GenAI_Memorisation_Open_Source_IEmanuilov_TMargoni-Preprint_Zenodo.pdf?include_deleted=0> accessed 31 July 2024; Julio Carvalho, 'The Stubborn Memory of Generative AI: Overfitting, Fair Use, and the AI Act' (*Kluwer Copyright Blog*, 8 April 2024) <https://copyrightblog.kluweriplaw.com/2024/04/08/the-stubborn-memory-of-generative-ai-overfitting-fair-use-and-the-ai-act/> accessed 31 July 2024.

[13] 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 11 October 2021; 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 31 July 2024.

[14] Cooper and Grimmelmann (n 12).

[15] 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).

[16] 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*).

and define "pastiche" as an autonomous concept of EU law, which could potentially apply to AI-generated outputs.[17]

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.[18].

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 the generative AI value chain.[19] This is also 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.[20] There are also critics of the copyright licensing of training data, arguing that it limits data quality and availability, increases costs, reduces competition, and slows innovation and productivity in GenAI, ultimately hampering economic growth.[21] Finally, 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 (on which, see some of deals struck by AI companies listed below).

---

[17] 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 31 July 2024; Piero Casanova, 'Permissible Pastiche in Pelham II: A Proposed Response' (*Kluwer Copyright Blog*, 11 April 2024) <https://copyrightblog.kluweriplaw.com/2024/04/11/permissible-pastiche-in-pelham-ii-a-proposed-response/> accessed 31 July 2024; Sabine Jacques, 'The Parody Exception: Revisiting the Case for a Distinct Pastiche Exception' (*Kluwer Copyright Blog*, 5 October 2023) <https://copyrightblog.kluweriplaw.com/2023/10/05/the-parody-exception-revisiting-the-case-for-a-distinct-pastiche-exception/> accessed 31 July 2024; Susan Bischoff, 'The Dawn of Pastiche: First Decision on New German Copyright Exception' (*Kluwer Copyright Blog*, 7 June 2023) <https://copyrightblog.kluweriplaw.com/2023/06/07/the-dawn-of-pastiche-first-decision-on-new-german-copyright-exception/> accessed 31 July 2024; 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) <https://freiheitsrechte.org/en/themen/demokratie/expert-opinion-on-pastiche> accessed 31 July 2024.

[18] 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 31 July 2024; Gabriele Cifrodelli and Lilian Edwards, 'Copyright and Generative AI: What Can We Learn from Model Terms and Conditions?' (*Kluwer Copyright Blog*, 24 July 2024) <https://copyrightblog.kluweriplaw.com/2024/07/24/copyright-and-generative-ai-what-can-we-learn-from-model-terms-and-conditions/> accessed 31 July 2024.

[19] 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.

[20] 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*, 23 July 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 31 July 2024.

[21] Bertin Martens, 'Economic Arguments in Favour of Reducing Copyright Protection for Generative AI Inputs and Outputs' <https://www.bruegel.org/working-paper/economic-arguments-favour-reducing-copyright-protection-generative-ai-inputs-and> accessed 31 July 2024.

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).[22] The Act is an extremely long and complex legislative text. Even for a highly motivated academic, it is a challenging and tedious read. The AI Act contains 108 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).

For our purposes, the most relevant provisions are found in Chapter V, on general purpose AI models, which contains the AI Act's copyright-relevant obligations.[23] 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 APIs or direct download, and can be modified or integrated into AI systems, which require additional components like 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. For our purposes, an example of the model is the generative pre-trained transformers or GPTs and the system is the chatbot or tool many use on a daily basis, such as ChatGPT, Midjourney, Dall-E, or Firefly).[24] As we will see below, this 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 a typical *example* 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.

---

[22] For some of the legislative story of the copyright provisions in the AI Act, see Quintais (n 2); Peukert (n 2).
[23] See also supporting recitals 104 to 109
[24] 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 .

The figure 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.[25]

The only relevant scope exclusion here is stated in Article 2(6).[26] This provision 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. This is 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 model" that is then commercialized[27]; or models based on scientific purpose/use GPAI models that are subsequently made available on the market.[28]
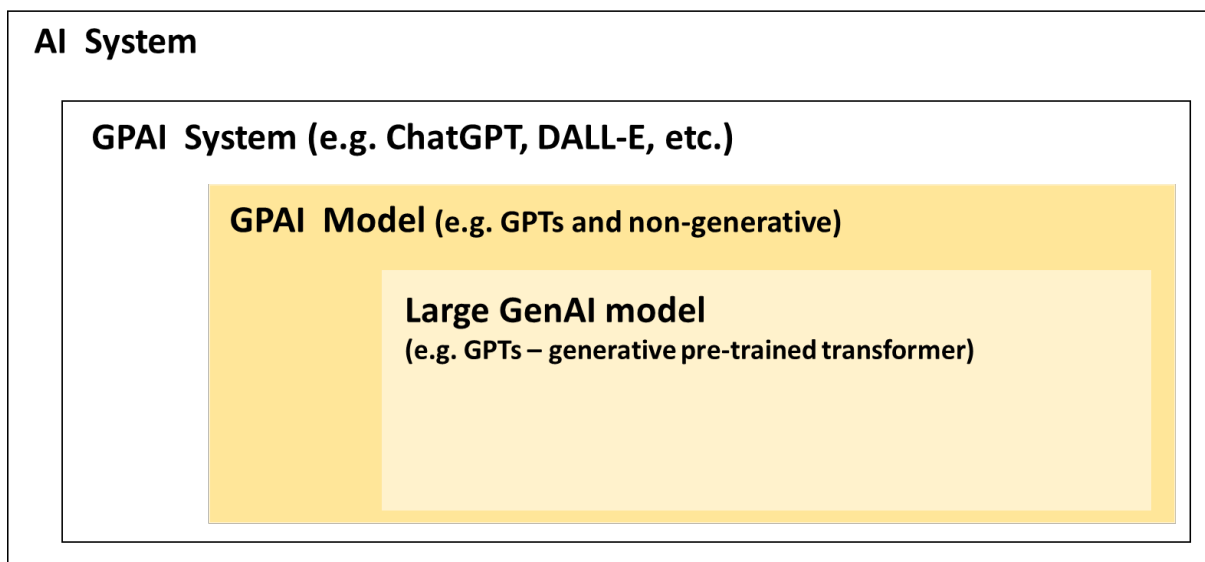
4. The AI Act vs EU copyright law interface

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[29] through the imposition of systemic compliance obligations on certain providers. 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

---

[25] Article 3(3) and (4)) AI Act.
[26] See also recitals 25 and 109 AI Act.
[27] Recital 109 AI Act.
[28] Recital 25 AI Act.
[29] See e.g. Article 1(1) AI Act.

private differences in nature lead to differences in enforcement and remedies.[30] As Peukert puts it, the copyright provisions of the AI Act are "a fusion of two different types of laws".[31]

In addition, like with the Digital Services Act (DSA)[32], the interaction between the AI Act and EU copyright law is far from being 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 2 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) 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 GenAI 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 TDM 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 GenAI models, and thus their lack of relevance for copyright purposes[33]. 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.

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).

---

[30] See infra at 4.6 and 4.7.

[31] Peukert (n 2).

[32] 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.

[33] 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 31 July 2024.

- 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 general-purpose AI model, according to a template provided by the AI Office.

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

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 GPAI *system* providers. To be sure, the two 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." 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".[34] This remains true despite the obligation to draw 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.[35]

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[36]– 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.[37] Going further down the rabbit hole, one assumes that if a generative AI system is embedded in an online content-sharing service provider, 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.

Why might this distinction between model and system matter? To be sure, if we consider the copyright relevant 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 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

---

[34] Peukert (n 2).
[35] Article 53(1)(b) AI Act.
[36] European Commission, 'Supervision of the Designated Very Large Online Platforms and Search Engines under DSA | Shaping Europe's Digital Future' (2024) <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses> accessed 31 July 2024.
[37] Recital 118 AI Act.

deployment of the AI system incorporating the model. In these scenarios, the model vs system distinction matters.

Second, because the copyright obligations in the AI Act are separate and independent from those in the copyright *acquis*[38] – 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.[39]

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 of sizes and models.[40] 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.[41]

Fourth, these obligations are to be monitored by the AI Office, a part of the European Commission.[42] 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 to me whether this means the same as an obligation of result to respect EU copyright law; it appears, rather, to be an obligation of means. Time will tell whether this obligation amounts to more than a formal requirement to have a policy in place, which would seem an unsatisfactory outcome and at odds with the spirit of the AI Act.

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, including downstream moderation of outputs. 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 (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-

---

[38] Recitals 108, 109 AI Act.
[39] See infra at 4.6 and 4.7.
[40] See Recital 104 and Article 53(2), not applying to paras (c) and (d)).
[41] Recital 109 AI Act.
[42] Recital 108 AI Act.

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.[43]

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".[44] 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 phase or only when content is publicly available online (e.g. during web scraping or harvesting)? 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 is at the center of the recent *Kneschke v LAION* proceedings.[45] It would seem however from Recital 18 CDSMD, at the very least, that an opt-out via terms and conditions of a website is only appropriate if it is also machine readable.[46]

In any case, a key aspect here is that the requirement of opt-out as designed in the CDSMD is not directed towards GPAI model providers. Rather, for content made publicly available online, it seems to be aimed primarily at website owners where protected content is hosted, which in many cases will not be the copyright holders or the GPAI model provider. The AI Act obligations, for their part, are squarely aimed at 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".[47] Current well-known examples of opt-out approaches include

---

[43] See infra at 4.6.

[44] See generally on the topic of opt-out, 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 25 June 2024.

[45] Keller, 'Machine Readable or Not?' (n 8).

[46] See Abbamonte (n 2).

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

Spawning AI's suite of tools[48] for rights holders (Have I Been Trained,[49] Kudurru,[50] or even a Browser Extension[51]), Google-Extended,[52], and OpenAI's announced Media Manager.[53] 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).[54] 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.[55] 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".[56]

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.[57]

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.[58]

---

[48] See https://spawning.ai/
[49] See https://spawning.ai/have-i-been-trained
[50] See https://kudurru.ai/
[51] See https://spawning.ai/browser-extension
[52] See, Google, Overview of Google crawlers and fetchers (user agents), https://developers.google.com/search/docs/crawling-indexing/overview-google-crawlers
[53] OpenAI, 'Our Approach to Data and AI' (*OpenAI*, 7 May 2024) <https://openai.com/index/approach-to-data-and-ai/> accessed 31 July 2024. For criticism, see Bernd Justin Jütte, 'Open AI's Vison for a Social Contract – of Things to Come…' (*Kluwer Copyright Blog*, 3 June 2024) <https://copyrightblog.kluweriplaw.com/2024/06/03/open-ais-vison-for-a-social-contract-of-things-to-come/> accessed 31 July 2024.
[54] 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 8 October 2023.
[55] as suggested e.g. by Arvind Narayanan and Sayash Kapoor, 'Generative AI's End-Run around Copyright Won't Be Resolved by the Courts' (20 March 2023) <https://www.aisnakeoil.com/p/generative-ais-end-run-around-copyright> accessed 31 July 2024.
[56] Article 50(2) AI Act.
[57] 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 31 July 2024.
[58] 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).

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." 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.

But I see two problems with this provision. 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".[59] In my view, the normative exhortation in this recital goes beyond the legal provision it supports, since it introduces extraterritoriality through the backdoor. This is because of the second problem: the territoriality principle of copyright law. In short, if I carry out the relevant TDM acts to pre-train and train the GPAI model outside the EU, then I'm not infringing Article 4 CDSM Directive if I only place the model on the market in the EU *post-training*. Naturally, if any of the TDM activities has a clear point of attachment with EU territory – most notably web scraping – then the model provider will have to respect EU copyright law, including the opt-out requirements.[60]

In sum, if the TDM leading up 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 CDSM Directive. As such, even if this recital is turned by national law into a binding obligation, its violation does not amount to copyright infringement. It can only be a violation of the AI Act, subject to the sanctions and penalties mentioned below. Even then, because this particular obligation refers back to the policies to respect copyright, it seems strange to impose a sanction on a provider that has in fact respected the applicable copyright rules.

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. That is to say, copyright interests should not be the only ones considered in shaping the contours of the obligation.[61]

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

---

[59] Maarten den Heijer, Teun van Os can den Abeelen and Antanina Maslyka, 'On the Use and Misuse of Recitals in European Union Law' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3445372> accessed 31 July 2024.
[60] For different interpretative options see Peukert (n 2); Senftleben (n 19).
[61] 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 25 June 2024.

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 (Recital 107).

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.[62]

From a copyright perspective, this obligation may help clarify two aspects that are essential to determine compliance with the TDM exception in art. 4 CDSMD. First, it may help establish whether relevant TDM activities took place – 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.[63]

Finally, as to the contents of such a summary, Warso, Gahntz & Keller advance a promising blueprint for a template.[64]

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 an European harmonized standard. 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 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 obligation 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". That is to say, it is only guaranteed that a Code of Practice will be developed to comply with the key transparency obligation explained above, but not the opt-out obligation. This may rsesult in the curious situation where no code of practice is developed for opt-outs until an harmonized standard emerges.

---

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

[63] On which topic, see Thomas Margoni, 'Saving Research: Lawful Access to Unlawful Sources under Art. 3 CDSM Directive?' (*Kluwer Copyright Blog*, 22 December 2023) <https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/> accessed 31 July 2024.

[64] Warso, Gahntz and Keller (n 61). (Disclosure: I was one of the academics providing feedback on this proposal).

This leads us then 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 approved code of practice, they 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 recently announced Media Manager for creators and content owners[65] meet the "policies" key obligation.

Interestingly, the AI launched on 30 July 2024 a multi-stakeholder consultation on trustworthy GPAI models under the AI Act; the first section of the questionnaire is on transparency and copyright-related provisions.[66]

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 be in compliance 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. Curiously, 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 both the AI Act and the CDSMD are subject to CJEU interpretation 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. In doing so, the Court will 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.[67] In other words, enforcement of these obligations rests squarely on public bodies, namely the Commission and the AI Office, which has 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

---

[65] OpenAI, 'Our Approach to Data and AI' (n 53).
[66] European Commission, 'AI Act: Have Your Say on Trustworthy General-Purpose AI | Shaping Europe's Digital Future' (30 July 2024) <https://digital-strategy.ec.europa.eu/en/consultations/ai-act-have-your-say-trustworthy-general-purpose-ai> accessed 31 July 2024.
[67] European Commission, 'European AI Office' (*European Commission*, 2024) <https://digital-strategy.ec.europa.eu/en/policies/ai-office> accessed 31 July 2024.

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, when the Commission finds that the provider acted intentionally or negligently. 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-a-vis GPAI model providers' obligations.

Although the potential fines are substantial, it is noted that 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 by the tsunami of agreements between providers and e.g. between providers and rights holders and aggregators. Examples include the recent Microsoft deal with Taylor & Francis[68] and the multiple deals OpenAI has in place with players such as TIME[69], the Atlantic,[70] the Financial Times,[71] Le Monde and Prisa Media[72], Axel Springer,[73], and The Associated Press[74].[75]

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.[76] Contrary to the US,

---

[68] Wellett Potter, 'An Academic Publisher Has Struck an AI Data Deal with Microsoft – without Their Authors' Knowledge' [2024] *The Conversation* <http://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203> accessed 31 July 2024.

[69] OpenAI, 'Strategic Content Partnership with TIME' (*OpenAI*, 27 June 2024) <https://openai.com/index/strategic-content-partnership-with-time/> accessed 31 July 2024.

[70] OpenAI, 'A Content and Product Partnership with The Atlantic' (*OpenAI*, 29 May 2024) <https://openai.com/index/enhancing-news-in-chatgpt-with-the-atlantic/> accessed 31 July 2024.

[71] OpenAI, 'We're Bringing the Financial Times' World-Class Journalism to ChatGPT' (*OpenAI*, 29 April 2024) <https://openai.com/index/content-partnership-with-financial-times/> accessed 31 July 2024.

[72] OpenAI, 'Global News Partnerships: Le Monde and Prisa Media' (*OpenAI*, 13 March 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 31 July 2024.

[73] Axel Springer, 'Axel Springer and OpenAI Partner to Deepen Beneficial Use of AI in Journalism' (*Axel Springer*, 13 December 2023) <https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism> accessed 31 July 2024.

[74] Associated Press, 'AP, Open AI Agree to Share Select News Content and Technology in New Collaboration' (*The Associated Press*, 13 July 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 31 July 2024.

[75] 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*, 30 July 2024) <https://authorsalliance.substack.com/p/what-happens-when-your-publisher> accessed 31 July 2024.

[76] 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).

where statutory damages may present the greatest risk for model providers[77], 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[78] could prove particularly 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. In judicial proceedings, their implementation will be challenging, not just from a technical perspective but also as regards the assessment of their proportionality vis-a-vis the type and level of infringement in the particular case.

5.   Conclusion

The emergence of generative AI has put stress on EU copyright law. Although Article 4 CDSMD is probably 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. 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 law. This difference results in varied obligations, targets, consequences for infringement, enforcement structures, and available remedies. The copyright obligations in the AI Act may positively influence the overall design of GPAI models with respect to copyright law and might even indirectly help the private enforcement of Article 4 CDSMD.

However, the most probable outcome—and one that we can already observe—is that commercial GPAI model providers are incentivized to enter licensing deals with large rights aggregators to ensure high-quality datasets and avoid legal exposure. So far, there is no evidence that these deals improve the remuneration of individual creators and artists, who were arguably the intended beneficiaries of the late introduction of copyright rules in the AI Act. 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. 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 is an important role for collective licensing and—more important still—for

---

[77] 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*, 11 June 2024) <https://cacm.acm.org/opinion/how-to-think-about-remedies-in-the-generative-ai-copyright-cases/> accessed 31 July 2024.

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

collective bargaining. This latter mechanism, largely underexplored in the copyright vs. AI discussion in Europe, has been[79] (and continues to be[80]) used with some success in the US by authors' and performers' unions; 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.

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, the introduction of an output-oriented AI levy system[81], or a human rights-based statutory license for machine learning purposes.[82] However, considering how recent the CDSMD is, none of these solutions seem likely.

***

---

[79] The Authors Guild, 'SAG-AFTRA Agreement Establishes Important Safeguards for Actors Around AI Use' (*The Authors Guild*, 17 January 2024) <https://authorsguild.org/news/sag-aftra-agreement-establishes-important-ai-safeguards/> accessed 31 July 2024.

[80] Sarah Parvini, 'Video Game Performers Will Go on Strike over Artificial Intelligence Concerns' (*AP News*, 25 July 2024) <https://apnews.com/article/sagaftra-video-game-performers-ai-strike-4f4c7d846040c24553dbc2604e5b6034> accessed 31 July 2024.

[81] Senftleben (n 19).

[82] Geiger and Iaia (n 19).