Copyright and Generative AI: Opinion

by **Séverine Dusollier, Martin Kretschmer, Thomas Margoni, Peter Mezei, João Pedro Quintais, Ole-Andreas Rognstad** *

Executive Summary: The ECS considers that the current development of generative artificial intelligence (AI), under the regulatory framework set up by the Directive on Copyright in the Digital Single Market (CDSM) of 2019 and the AI Act of 2024 (Regulation (EU) 2024/1689), leaves legal uncertainties and several open questions. The following issues require, in the view of the ECS, urgent consideration by the European Union:

1. The determination of the scope of the text and data mining (TDM) exception: the exception enacted in Arts. 3 and 4 of the CDSM Directive at a time when the Generative AI development could not have been fully anticipated, can be interpreted as covering some operations of training of a Generative AI model, but certainly not all aspects or stages of the life cycle of AI models and systems, from curating a dataset for training to the generation of an image, text or other media, by users. The exact scope of the TDM exception, and hence the copyright status of acts carried out at each stage of development and operation of Generative AI models and systems, should be further studied and analysed. That would require a decision as to whether acts of reproduction or public communication occur and which actors are liable for such acts. Under such an assessment, the possibility of commercial use of models trained for scientific research and the effect of the exercise of the opt-out provided by Art. 4 CDSM Directive, on the availability of lawfully accessible sources for the research exception provided by Art. 3 CDSM Directive, merit particular attention.

2. The content of the obligation under Art. 53(1)(c) of the AI Act related to the reservations of rights: in particular, the technologies that can be used to express the opt-out should be identified and regularly reviewed; the rightholders entitled to opt-out and the opt-out modalities, including the timing and the location, should be clarified.

3. The scope and modalities of the transparency obligation laid down by Art. 53(1)(d) of the AI Act: in particular, the relevant information to be included in the summary and the impact of the transparency obligation on the assessment of the lawful access criterion contained in Arts. 3 and 4 CDSM Directive should be clarified.

4. The privileges for research and for open source models: the importance of research and the key role of open source data and software in the field of AI should guide the interpretation of the CDSM Directive and the AI Act. This would lead to needed clarification of some of their provisions, with the objective of preserving the fundamental rights of research, academic freedom and education. The uncertainties raised by the Hamburg court decision in the LAION case, as to the interface between Art. 3 and Art. 4 of the CDSM Directive, should particularly be addressed in order to avoid general purpose AI (GPAI) model providers relying on training for the purposes of research, hereby escaping the more restrictive frame of the exception of Art. 4.

5. The articulation between the CDSM Directive and the AI Act: the CDSM directive is a private law instrument organizing a protection of private rights on a territorial basis, whereas the AI Act is a public law that regulates the safety of AI products, as a condition for importation and use in the EU. That raises several issues in the articulation of both legislative texts, notably the territorial scope of the obligations imposed, the entities covered by the different obligations, the effect of the AI Office's voluntary Code of Practice, the distinct modes of enforcement of the obligation laid down by the CDSM Directive and by the AI Act. These points should be clarified.

6. The fair remuneration of authors and performers for all acts of exploitation of their works and performances occurring in the life cycle of Generative AI models and systems (including when an opt-out from the application of Art. 4 CDSM Directive has been exercised and when their works or performances are included in a dataset that has been licensed to an AI provider) needs to be reaffirmed as a fundamental principle of the EU acquis. The Commission should look at the best ways to ensure such a remuneration, including remuneration rights or other compensation mechanisms, in concert with Member States.

*   The European Copyright Society (ECS) was founded in 2012 with the aim of creating a platform for critical and independent scholarly thinking on European Copyright Law and policy. Its members are scholars and academics from various countries of Europe, seeking to articulate and promote their views of the overall public interest on all topics in the field of authors rights, neighbouring rights and related matters. The ECS is neither funded nor instructed by any particular stakeholders. Its Opinions represent the independent views of a majority of ECS members.

Background

Before the advent and public availability of generative AI tools such as ChatGPT, Stable Diffusion, MidJourney Dall-E, GitHub or Udio, the Directive (EU) 2019/790 of 17 April 2019 on copyright and related rights in the Directive on Copyright in the Digital Single Market (CDSM Directive) enacted two exceptions to copyright and related rights to allow for text and data mining (TDM) of protected subject matter: one for purposes of scientific research, the other for any other purpose. In that latter case, the rightholders are entitled to reserve the right to authorise such TDM by opting out of the application of the exception. In 2024, in the context of a growing concern that generative AI tools could produce texts, images, music or films and impact copyright protection and remuneration of creators and artists, the Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act), in its final stage of negotiation, included obligations to providers of general-purpose AI (hereafter GPAI) models to provide transparency as to the datasets used for training their models and to put in place a standard policy for the exercise of opt-out by copyright and related rights owners.

In parallel with the growth of litigation in the US and Europe, controllers of aggregate copyright works (such as news publishers and outlets, stock images companies, or other types of content) are striking deals with technology firms about (often exclusive) access for AI model training. At the same time, authors, artists, performers are receiving new contract types from publishers, producers and collective management organisations (CMOs). These compete for assignments or clarifications about rights to train, which the respective intermediaries aim to license on to technology companies.

Due to the discussions surrounding the adequate manifestation of the CDSM Directive opt-out provision and other challenges, including in other jurisdictions, in effect the AI training space is already moving to licensing as a default.

Without contesting what has been achieved by the CDSM Directive and by the AI Act, the ECS considers that the rapid development of generative AI technology associated with the emergence of a licensing market for specific datasets, highlight some remaining uncertainties and bring new challenges that require EU intervention.

The ECS is also following with attention the drafting of the General-Purpose AI Code of Practice initiated by the Working Groups set up by the EU AI Office,

particularly in relation to copyright. However, we would like to point out several pending questions, issues and uncertainties related to the combined application of both legislative texts to generative AI models and systems and copyright protection. Those remaining issues and uncertainties are of great policy relevance and are critical to innovation and to the sustainability of a distinct European creative sphere. The EU copyright *acquis* is founded in the fundamental rights framework established by the Treaties and the Charter of Fundamental Rights of the EU (Charter). The AI Act, which necessarily operates within the same framework, is perhaps even more explicit in its fundamental rights enabling objectives. Art. 1 of the AI Act puts this clearly when it states that its purpose is to "... promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection ...".

As is well established by the case law of the CJEU, neither of the fundamental rights established in the Charter is absolute or prevails over the others, but they are in a constant need of balancing and dialogic conversation. The ECS believes that in the balancing exercise needed to address the many possible tensions across the multifaceted actors in the AI life cycle the following elements should operate as guiding principles:

- The interests of human authors and performers;

- The interests of users and of the wider public, anchored in the fundamental rights framework established by the Treaties and the Charter, as reminded by Art. 1 of the AI Act;

- The enhancement of research and innovation.

1. The application of the text-and-data-mining exception to generative AI operations

Although the issue has been disputed, the ECS holds the view that the TDM exceptions in Arts. 3 and 4 CDSM Directive are applicable to the development of generative AI models (a type of GPAI model as per the AI Act), albeit not necessarily covering all aspects of it. While it is not *prima facie* obvious that the reproduction right in Art. 2 of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society (Infosoc Directive) or the extraction right of Art. 7 of Directive 96/9/EC (Database Directive) apply to any case of training of generative AI models, Recital 105 of the AI Act nevertheless presupposes that "any use of copyright protected

content requires the authorisation of the rightholder concerned unless relevant copyright exceptions and limitations apply". At the same time, this Recital as well as substantive provisions of the AI Act (e.g. Art. 53(1)c)) are based on the premise that the TDM provisions of the CDSM Directive are applicable to the development of generative AI systems or GPAI models. The broad phrasing of the definition of TDM in Art. 2(2) of the Directive ("any automated analytical technique aimed at *analysing text and data* in digital form in order to **_generate_** information which includes but is *not limited to* patterns, trends and correlations"; emphasis added here) supports this position.

However, although generally relevant, the TDM provisions of the CDSM Directive should not be considered as resolving all the problems concerning the use of works protected by copyright and other subject matter. The provisions of Arts. 3 and 4 of the CDSM Directive do not necessarily apply to all aspects or stages of the life cycle of a GPAI model or a generative AI system. Rather, these TDM exceptions cover different (but not necessarily all) aspects of the training stage of GPAI models, to the extent that those activities qualify as: (i) acts of TDM under the broad definition in Art. 2 of the CDSM Directive, and (ii) reproductions of protected subject matter of the type mentioned in Art. 3 and 4 of the CDSM Directive.

This raises two important questions. First, does the concept of TDM and its exceptions cover all activities taking place leading up to and including the training stage of a GPAI model? Second, is the TDM regime relevant for acts taking place once a model is trained and when outputs are generated?

Regarding the first question, the AI Act takes a clear position on the copyright-relevant nature of TDM, as already stated above. Recital 105 also mentions the rights reservation mechanism in Art. 4(3), noting that, where applicable, GPAI model providers must abide by this mechanism if they want to conduct TDM on those materials, namely by implementing the principles agreed upon by the Code of Practice to be established by the EU AI Office.

From the perspective of EU law, therefore, carrying out TDM on copyright-protected content appears in most cases to amount to reproducing a work.[1] As such, TDM requires authorization from the rightholder, or it must benefit from a copyright exception, such as those in Art. 3 and 4 of the CDSM Directive. The question that arises is whether all copyright-relevant reproductions and extractions involved in the training and development of an AI model qualify as TDM. What appears clear is that the TDM exception does not cover subsequent acts of communication to the public or the making available of TDM results. Indeed, the scope of the TDM exceptions covers only acts of reproduction and extraction. Furthermore, Art. 3(2) and 4(2) of the Directive put clear boundaries on the subsequent uses of copies of works or other subject matter made pursuant to the TDM exceptions.

Regarding the post-training operations, it should be noted that the TDM definition and the TDM exceptions do not apply to any acts taking place at a stage following to training the model. This means that they do not cover the integration of a trained GPAI model into an "AI system," its "placing on the market," "making available on the market," or its "putting into service" in the EU. They also do not cover the generation of outputs by an AI model or system.

All these activities may be relevant for copyright purposes, as they may involve restricted acts and subsequent copyright infringement. There is significant legal uncertainty about the copyright status of acts that have been labelled as memorization at the model level, as well as regurgitation, extraction, and reconstruction at the output generation stage. The integration of a dataset constituted or of a model trained under the research-related TDM exception (art. 3 of the CDSM Directive) in a GPAI model or system made available for commercial purposes is another issue that has been recently dealt with by the Hamburg District Court (the LAION case)[2] in a manner that raises many questions. These issues deserve further research and clarification.

As a result, the TDM exception and its assessment should be considered separately from the commercial exploitation, effects, or harms to creators stemming from generative AI outputs. In other words, such commercial exploitation and competition with or substitution for human (non-AI-assisted or generated) creations are not relevant – as a matter of law – to the assessment of the exception in Art. 4 of the CDSM Directive. This also has consequences for the assessment of the exception under the three-step test, as the qualification of a conflict with the normal exploitation and the assessment of unreasonable prejudice to rightholders must be considered in the context of TDM related to the training of an AI model, rather than in relation to the exploitation that takes place once the model is

---

1   This policy choice of including any technical, even if fugitive, fixation of a work within the scope of reproduction right, made by the EU lawmaker as early as the 1991 directive on computer programs, could have been different and remains challenged by several copyright scholars, including some signatories of the present opinion.

2   Landgericht Hamburg, 27 September 2024, AS. 310 O 227/23.

trained and placed on the market. By contrast, the commercial exploitation that takes place during the training stage – e.g., licensing of datasets by rightholders for third parties to carry out TDM – might be relevant to the assessment of Art. 4 of the CDSM Directive. This is a point of legal interpretation of EU law, rather than a normative pronouncement on its desirability.

2. The content of the obligation under Art. 53(1)(c) of the AI Act

According to this rule, GPAI model providers must *put in place a policy to comply with EU copyright law* in particular to identify and comply with, including through state of the art technologies, the reservations of rights (i.e. "opt-out") expressed pursuant to Art. 4(3) of the CDSM Directive.

This provision includes two main prongs: on the one hand, the requirement to 'put in place' a policy document; and, on the other hand, to identify and comply with opt-out mechanisms, that is, in essence, to guarantee the compliance with the CDSM Directive.

As regard the first prong, 'putting in place' a policy shall not only mean *drawing up* such a document, but GPAI model providers shall also *keep such policy up-to-date*, they shall also *implement* their commitments per the policy document, and, finally, to *publish* the policy document. The latter shall be understood in a broader sense: GPAI model providers shall provide access to the policy document to the general public, rather than solely to the AI Office. This is evident from the language of the AI Act itself. Art. 53(1)(a) of the AI Act introduces a limited publication obligation ('upon request, to the AI Office and the national competent authorities'); whereas Art. 53(1)(c) does not include any such limitation.

As regards the second prong, GPAI model providers' policy, in line with *effet utile*, only if capable to guarantee that rightholders can effectively opt-out their contents from the training of GPAI models.

Based on that, the Commission and the AI Office, and particularly the Working Group on transparency and copyright-related rules, have already started to work on a Code of Practice to provide guidance on (a) the *scope and modalities of the said policy requirement*; (b) the modalities and methods of the opt-out mechanism that will be considered *compliant with Art. 53(4) AI Act*.

In that process, the *effect of the compliance of GPAI model providers with the obligation under Art. 53(1)(c) AI Act* on the consideration of whether they are compliant with the rights reservation rule under

Art. 4(3) CDSM Directive, should be ascertained and a special clarification is needed regarding various sub-topics. First, as provided for by Art. 56(8) of the AI Act, technologies to be used for the expression of rights reservation need to be regularly reviewed, in order to avoid the danger that a specific technological solution becomes mandatory and to ensure instead that all state-of-the-art solutions might be deployed in practice. Second, rightholders entitled to opt-out and the opt-out modalities should be expressly determined; an issue that has special importance in light of the numerous alternatives for opt-outs (developed by GPAI model providers and/or independent third parties) and the growing number of "press-release-like" reservation of rights by CMOs or licensees, e.g. publishing houses. Third, the timing of the reservation of rights should be discussed; that is, whether opt-outs preceding or following the mining of text or data are compliant with the *acquis*. Finally, the location of the expression of the reservation should be clarified; that is, whether opt-out at the source-level where the protected subject matter is stored or from where it has been made lawfully accessible and/or at the work-level, that is, via the developers' website/reservation mechanism, are covered by the *acquis*.

3. The scope and modalities of the transparency obligation laid down by the Art. 53(1)(d) of the AI Act

These rules require GPAI model providers to draw up and make publicly available a sufficiently detailed summary about the content used for training of the GPAI model (including of the generative type), according to a template provided by the AI Office.

First, our arguments expressed in the previous point on 'drawing up', 'publish' and making the relevant document 'meaningful' apply mutatis mutandis under Art. 53(1)(d) of the AI Act. Similarly, under *effet utile*, the summary shall include relevant information about how and when the providers respected opt-outs required by Art. 4(3) of the CDSM Directive.

From a copyright perspective, it is also crucial that the Commission and the AI Office clarify *how this requirement of the AI Act influences the assessment of the lawful access criterion* (or even criteria) underpinning Arts. 3 and 4 CDSM Directive and what exact information GPAI model developers shall disclose as regards such access to training data.

4. The privileges for research and for open source models

The AI Act acknowledges the importance of research in the field of AI as well as the use of AI in research activities. It therefore establishes that it 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 (Article 2(6)). The AI Act also does not apply 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 (Article 2(8)). These are important provisions. Yet, given the very strict definition of research, their practical effect, particularly in the context of public-private partnerships in research, remains to be ascertained.

In a similar vein, the AI Act recognizes that software and data, including models, released under a free and open-source license can contribute to research and innovation in the market and can provide significant growth opportunities for the Union's economy. Accordingly, the AI Act does not apply to AI systems released under such licenses, unless they are placed on the market or put into service as prohibited AI or as high-risk AI systems or as certain AI systems subject to specific transparency obligation (Art. 50). GPAI models under free and open-source licenses are excluded only from the provisions of Art. 53(1)(a) and (b), but must comply with those under letter c (the Policy) and letter d (the Summary). Essentially, the documentary obligations of Art. 53(1)(a) and (b), together with the exclusion for certain third parties in relation to the high-risk AI value chain (Art. 25(4)) are the only actual exemptions favoring free and open-source AI. Considering the restrictive definition adopted in the AI Act that excludes any form of monetization – a considerable deviation from the generally accepted definitions of free and open source software – the real effect of the provision, similarly to the case of research, remains unclear.

A very specific issue was highlighted by the LAION case, brought before the Hamburg district court: the potential use by commercial players of datasets mined on the basis of Art. 3 CDSM Directive. Whereas this aspect would deserve a dedicated treatment, the analysis needs to take into account, as argued above, that both the copyright *acquis* and the AI Act are grounded in the fundamental rights framework established by the Treaties and the Charter. Scientific research, academic freedom and the right to education are central in this framework and their preservation must be ensured. In the specific case of LAION, the dataset prepared did not contain the actual works needed for the successive phase of model training, but only information about their location. A GPAI model provider interested in exploiting this "preselection" would need a proper legal basis to access those sources. This legal basis would likely be Art. 4 of the CDSM Directive or a contractual agreement, in case the opt-out provided for in Art. 4(3) has been exercised. In the opinion of the ECS this approach, already logically following from the regulatory framework put in place by the interface between the Arts. 3 and 4 of the CDSM Directive and Art. 53 of the AI Act, represents a proportionate balance in the protection of the different fundamental rights at stake.

5. The relationship between the AI Act (a Regulation) and the CDSM Directive with respect to the enforcement of the copyright-related provisions

Recital 108 clarifies that the AI Act does not affect the enforcement of copyright rules as provided for under Union law; several recitals and provisions mention that the AI Act is both without prejudice to Union copyright law or meant to assist in compliance with EU copyright law.

Copyright law is an area of private law where civil enforcement is left to the owners of copyright and related rights. In part, the relationship between the AI Act and copyright law is just a clarification and assertion of such existing private interests of legal subjects, leaving enforcement within the national regimes of Member States, harmonised by the TDM provisions of the CDSM Directive.

However, an important (and entirely new) set of obligations in the AI Act need to be understood as meta-laws at the EU level. As Peukert suggests, they resemble "horizontal meta-obligations of hosting service and search engine providers under the Digital Services Act (DSA) who also have to put in place various mechanisms to act on or prevent the presence or findability of illegal content".[3]

The obligations of the AI Act about transparency and compliance with opt-out provisions of Art. 4(3) CDSM Directive imposed on the GPAI models providers are presented as if they have an *extraterritorial* effect, and could apply to the training of models outside of the EU. Breaches of obligations lead potentially to administrative fines (up to 3% of the annual total worldwide turnover or EUR 15 000 000, whichever is higher), i.e. a public law remedy rather than private enforcement.

3   A. Peukert (2024) Copyright in the Artificial Intelligence Act – A Primer, GRUR International, 73(6), 2024, 497–509, at p. 502.

The core of the AI Act, before the late introduction of rules covering foundation models and generative AI as GPAI in Chapter 5, introduced extraterritorial implications via the concept of the *AI value chain.* Under Art. 25 of the AI Act (Chapter 3, High-Risk AI Systems, Responsibilities along the AI value chain), the prohibitions and obligations for high-risk AI systems apply to any "distributor, importer, deployer or other third-party". However, these do not apply to development activity that takes place before the release and they do not include copyright obligations.

With respect to the copyright-related meta-obligations under Art. 53 of the AI Act, extraterritorial application relies on a supporting Recital 106 that demands compliance with EU law on copyright and related rights "regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place." This recital arguably goes beyond the legal provision it supports, potentially dislodging the territoriality principle of copyright law,[4] under which the provisions of EU copyright law do not apply outside its jurisdiction.

Furthermore, the entities carrying out copyright-relevant acts such as TDM-type reproductions, may not be model providers. That means their activities, such as those by Common Crawl (for web scraping) or LAION (for dataset preparation), will not fall under the GPAI chapter of the AI Act at all.

The European Commission is side-stepping the issue with the AI Office's voluntary Code of Practice under the instruction of Art. 56(1) of the AI Act. The extraterritorial effect of the (draft) Code's provision is indirectly obtained by wording that it applies "to all phases of the development of a general-purpose model, including data collection, training, testing and placing on the market" (Measure 2.1: Draw up and implement an internal copyright policy, second draft published 19 December 2024, Rules related to Copyright, AI Act Art. 53(1)(c)). Consequently, the life cycle approach of the Code of Practice will enable providers to demonstrate compliance with the AI Act, suggesting a complex form of voluntary extraterritoriality.

The introduction of value chain and life cycle concepts, combined with a mix of private and public law enforcement is new to the copyright sphere and needs to be thought through carefully. While the AI Act currently does not envisage private enforcement (e.g. a claim for damages from copyright and related rights owners), it may be fruitful to explore analogies

with competition law where findings of anti-competitive behaviour may lead to private action for damages.

6. The fair remuneration of authors and performers

Finally, a market is already developing for licensing of copyrighted works and other protected subject-matter, particularly to provide high-quality datasets for training generative AI models and systems (as demonstrated by recent examples of licensing partnerships between AI operators and press publishers, news outlets or images databases producers) and will continue to develop. Therefore, the question of a fair remuneration of authors and performers in compliance with the fundamental principle laid down by Art. 18 of the CDSM Directive needs to be addressed. The following principles should in our view apply to ensure that authors and performers are associated with any exploitation of their works and performances in generative AI operation:

- Art. 18 of the CDSM Directive mandates, as a general principle, that authors and performers receive an appropriate and proportionate remuneration for acts of exploitation of their works and performances in all relevant stages of operation of generative AI models and systems (from training to post-training commercial exploitation of generative AI models, as well as exploitation of generated content similar to their works or performances).

- When their works or performances are part of a collection of works that is specifically licensed to a generative AI model provider as a training dataset, the producer of such a collection, database or news publications needs to ensure an appropriate and proportionate remuneration to authors and performers of content included in the licensed dataset.

- When, after having opted out from the application of the TDM exception, under the conditions laid down by Art. 4 CDSM Directive and Art. 53(1)(d) of the AI Act, rightholders enter into licensing agreements to authorise TDM of works and other protected subject matter by generative AI model providers, some appropriate and proportionate remuneration should be provided to authors and performers when they have transferred or licensed their rights to such rightholder. Since remuneration in such a case of training on massive numbers of works and performances might be rather minimal for authors and performers or difficult

---

4 João Pedro Quintais, The AI Act, Copyright and extraterritoriality, Kluwer Copyright Blog, 28 November 2024.

to determine, it would be useful to investigate and identify the legal options left to Member States or adopted at the EU level to organise some other forms of appropriate compensation (such as a residual remuneration right or collective remuneration models existing in several Member States, or, beyond the copyright regime, other compensation mechanisms such as a financial contribution to cultural funds/activities or to the impacted creative sectors).

Disclaimer. *ECS member Prof. Alexander Peukert is currently chairing the sub-working group on the copyright-related provisions of the EU General Purpose AI Code of Practice under the AI Act. He did not participate in the drafting of this Opinion and takes no position on its contents.*