

Kluwer Copyright Blog

Copyright, the AI Act and extraterritoriality

João Pedro Quintais (Institute for Information Law (IViR)) · Thursday, November 28th, 2024

Introduction

The interaction between the AI Act (Regulation 2024/1689) and the exceptions for text and data mining (TDM) in the CDSM Directive is one of the most important topics in EU copyright law today. One particularly controversial point of intersection is the AI Act's attempt, through recital 106, to give extraterritorial effect to its copyright-related provisions. This blog post addresses that specific topic. Readers interested in this and other aspects of the interaction between these legislative instruments are invited to read the detailed analyses e.g. in [Quintais 2024](#) and [Peukert 2024a](#).



Generated with OpenAI's DALL-E 3 model based on a prompt corresponding to the wording of Recital 106 AI Act

The AI Act is an extremely long and complex legislative text, with 180 recitals, 113 articles and 13 annexes. Structurally, it is divided into 13 chapters. From the copyright perspective, the most relevant provisions are found in Chapter V, on general-purpose AI (GPAI) models, which contains the AI Act's copyright-relevant obligations.

The AI Act establishes two key copyright-related obligations that apply solely to GPAI *model* providers in Article 53(1)(c) and (d). That is to say, these provisions do not directly apply to upstream players (e.g. LAION when providing datasets) or downstream players, like AI systems providers or deployers (as a general rule).

The copyright rules in the AI Act are intended to interface with the TDM exceptions in the

CDSMD. In essence, Recital 105 AI Act states that if you carry out 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, like those in Articles 3 (scientific research TDM) and 4 (all purpose TDM, including of a commercial nature) of the Directive.

The first copyright obligation in the AI Act is found in Article 53(1)(c), which states that GPAI model 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.

The second obligation, in Article 53(1)(d), states that 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.

The extraterritoriality provision in the AI Act refers explicitly to the first obligation on “policies to respect copyright”, especially the opt-out.

Policies, opt-out and extraterritorial effect

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

Abbamonte labels this as a “market entry requirement” that is derived from the general rule in Article 2(1)(a) AI Act, according to which the Regulation applies *inter alia* to providers placing on the market GPAI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country. This is essentially also the argument advanced by [Stieper and Denger 2024](#), who consider it sufficient to conclude that as a result EU copyright directives “acquire direct effect vis-à-vis private legal entities outside the EU”. Color me skeptical.

Assuming the “copyright policies” obligation means more than a formal requirement to put a policy in place, there are at least two problems with this extraterritorial provision from the perspective of legal interpretation.

First, the provision is contained in a recital. Recitals are not binding and their primary function in EU law is “to explain the essential objective pursued by the legislative act” ([den Heijer et al. 2019](#)). As the CJEU has consistently stated, recitals cannot directly create rights or duties (e.g. [C-136/04](#), para. 23; [C-134/08](#), para 19). Therefore, recitals cannot be of a norm-setting character ([EC Legal Service 2015](#)). In my view, recital 106 does not aim to clarify the application of the general market location approach of the AI Act to substantive copyright rules. Rather, the recital is intended to support the interpretation of the provision containing the obligation to which it explicitly refers, located in Article 53(1)(c). In doing so, it establishes an additional norm on

extraterritoriality that extends beyond the enacting provision it references. In other words, the normative exhortation in recital 106 exceeds the legal provision it supports, as it broadens the territorial scope of the copyright policy obligation—and by extension, substantive copyright rules—particularly the opt-out obligation.

This leads us to the second and related problem: the territoriality principle of copyright law and the application of the rule of *lex loci protectionis* (the law of the country for which protection is claimed) to TDM activities (van Eechoud 2024). The territoriality principle binds the grant and effect of copyright to the territory of the state where protection is conferred to a work. Protection is granted on a territory-by-territory basis, with the attached exclusive rights following the same logic. It is therefore crucial to localize the relevant restricted act, for instance the reproduction of a work for TDM purposes. In the EU, Article 8 of the [Rome II Regulation](#) deals with conflict of laws regarding non-contractual obligations. This provision clarifies that the *lex loci protectionis* applies to infringement of copyright, covering both the requirements and scope of protection (Stieper and Denger 2024).

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

Peukert 2024b calls this a “minimalist” solution, as opposed to the “maximalist” approach of postulating the extraterritorial application of the AI Act in disregard of the principle of copyright territoriality (pp.9-12). He then advances an “intermediate solution”, which consists of “making the application of Art. 53(1)(c) AI Act dependent on whether the model provider scraped websites hosted on servers located in the EU.” (pp.11-12).

Arguably, however, this would already result from the application of the normal (minimalist) approach in accordance with the principles of territoriality and *lex loci protectionis*. In my view, if any of the TDM activities has a clear point of attachment with EU territory—most notably web scraping—then the model provider should have to respect EU copyright law, including the opt-out requirement (for a range of possible interpretations, see Peukert 2024a and Senftleben 2023).

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

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

a broader process connected to the EU.

But this interpretation is difficult to reconcile with the practice and rules applicable to TDM. There is a clear factual and legal distinction between (1) TDM activities required to train and build a model, and (2) the making available of that model on the EU market. These activities are treated differently under EU copyright law, where it is not possible to conflate the legal regime of TDM – which applies to rights of reproduction and extraction – and the act of making available a trained GPAI model. Furthermore, these activities are also distinct in the logic of the AI Act, which demarcates the training of a GPAI model from the subsequent integration of the trained model in an “AI system”, its “placing on the market”, “making available on the market”, and “putting into service” (see *inter alia* definitions in Article 3 (9)-(11), (63), (66) AI Act). In my view, no amount of functional interpretation can cross this particular interpretive Rubicon.

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

Code of Practice points the way forward?

One possible solution to this problem is currently being explored in the [First Draft of the General-Purpose AI Code of Practice](#) published in November 2024.

Article 56 AI Act regulates such codes of practice, which are to be drawn up with the AI Office acting as a facilitator. Although codes of practice are in principle “soft law”, sometimes characterized as “meta regulation” (Bygrave and Schmidt 2024) the Commission may, by way of an implementing act, approve a code of practice and give it general validity within the Union (Article 56(6)). Relatedly, the AI Act establishes that codes of practice must be ready at the latest by 2 May 2025; if that is not the case, the Commission may impose, through implementing acts, common rules for *inter alia* the copyright-related obligations in Article 53 (Article 56(6)).

The proposed draft Code of Practice deals with copyright issues in pages 14 to 16. The logic of the Code is to identify a high level measure, followed by concrete sub-measures and specific key performance indicators (KPIs) for such sub-measures. At this stage, concrete KPIs are largely absent. The following table provides an overview of the copyright sections.

Key copyright obligation	Measure	Sub-measure
Article 53(1)(c) - Policies	3. Put in place a copyright policy	3.1. Draw up and implement a copyright policy
		3.2. Upstream copyright compliance
		3.3. Downstream copyright compliance
	4. Compliance with the limits of the TDM exception	4.1. Respect Robots.txt
		4.2. No effect on findability
		4.3. Best efforts regarding other appropriate means
		4.4. Commitment to collaborative development of rights reservations' standards
		4.5. No crawling of piracy websites
	5. Transparency	5.1. Public information about rights reservation compliance
		5.2. Crawler name and robots.txt features
		5.3. Single point of contact and complaint handling
		5.4. Documentation of data sources and authorisations

As a preliminary remark, it is noteworthy that Measure 5 on transparency is linked to compliance with the “policies obligation” under Article 53(1)(c), rather than the actual transparency obligation outlined in subparagraph (d). Notably, Article 53(1)(d) is absent from this draft of the Code of Practice, likely because the drafters have not yet proposed a transparency template to meet this obligation.

Back to the topic of extraterritoriality. Sub-measure 3.1. indicates that the “policies obligation” should be understood in the context of the AI Act’s coverage of the “entire lifecycle” (Recital 109) of GPAI models.

Instead of mandated extraterritoriality, the Code appears to suggest a form of voluntary extraterritoriality, where the GPAI model provider would agree to only make available a model on the EU market where that model has complied with EU copyright law throughout its value chain or lifecycle.

In that context, for example, sub-measure 3.2 tries to reinforce this logic through due diligence and contractual operations, thereby tackling a major loophole of the Act vis-a-vis upstream providers (e.g, LAION, on which see [here](#), [here](#) and [here](#)). It states that GPAI model providers will undertake a reasonable copyright due diligence before entering into a contract with third parties concerning the use of data sets for the development of GPAI models, including whether these third parties respected TDM opt-outs under Article 4(3) CDSMD. Naturally, this commitment would also apply where such TDM activities took place outside the EU.

Moreover, there is a clear intent of not restricting compliance with any of the sub-measures 4.1 to 4.5 to TDM activities taking place in the EU. This is nowhere clearer than under sub-measure 4.5, where Signatories commit to take reasonable measures to exclude pirated sources from their crawling activities, including those listed in piracy watch lists published by relevant public authorities in the jurisdictions where GPAI Model providers are established, which presumably includes also non-EU jurisdictions.

In my view, this approach to voluntary extraterritoriality through meta-regulation with strong policy and market incentives is consistent with EU copyright law and provides a more promising avenue to ensure compliance with the “copyright policies” obligation.

This post is based on a section of the author’s working paper “Generative AI, Copyright and the AI Act (v.2) (November 01, 2024)” available at: <https://ssrn.com/abstract=4912701>. The author would like to thank Martin Kretschmer, Thomas Margoni, Séverine Dusollier and Alina Trapova for comments and suggestions on earlier drafts of the text, as well as for discussions on the topic.

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please [subscribe here](#).

A graphic for a survey report. It features a dark background with a circular inset showing a gavel on a glowing digital circuit board. The text is white and blue. A blue button with a white arrow points to the right.

2024 Future Ready Lawyer Survey Report

Legal innovation: Seizing the future or falling behind?

Download your free copy →

 Wolters Kluwer

 Future Ready

LAWYER

This entry was posted on Thursday, November 28th, 2024 at 8:18 am and is filed under [Artificial Intelligence \(AI\)](#), [CDSM Directive](#), [Digital Single Market](#), [European Union](#), [Legislative process](#), [Text and Data Mining \(TDM\)](#)

You can follow any responses to this entry through the [Comments \(RSS\) feed](#). You can leave a response, or [trackback](#) from your own site.

