STRATEGY

Policy Brief
November 2023

Author
Anthi Akritidou
Senior Public Affairs Manager
at EU Strategy

The AI Act and the Copyright Odyssey

Summary

The much-anticipated final trilogue on the AI Act has intensified discussions on the intersection of AI and copyright, particularly the Text and Data Mining (TDM) exception. Originating from amendments introduced by the European Parliament (as part of transparency obligations for providers of foundation models to make publicly available a summary of the copyright-protected content used to train the AI algorithms), debates emerged around the topic with organisations representing rightholders supporting disclosure and others questioning its practicality. The concept of foundation models, embraced by the European Parliament, appeared to have reached a consensus by the co-legislators in the latest trilogue held in October. However, concerns raised from large Member States such as France, Germany, and Italy in a subsequent technical meeting led to a dispute. The European Commission's intervention proposes a compromise, marking the first official reference to reinforcing the TDM exception, as laid down in the recent Copyright Directive. Simultaneously, the TDM exception itself faces challenges in effective enforcement due to the lack of harmonised standards. And so poses the question as to whether the existing copyright regime is suitable for the AI age. Key figures in the debate advocate for a copyright revision signalling potential developments in the upcoming mandates of the European Commission and Parliament.

State of play of the AI Act with regard to copyright provisions

As we are heading towards the final -hopefully- trilogue on the AI Act scheduled for the 6th of December, all eyes are focused on the final pending issues to be agreed upon by the co-legislators. Of particular interest is the intersection of the proposed AI rule-book with the copyright framework in place, notably the Text and Data Mining (TDM) exception laid down in the recent Copyright in the Digital Single Market Directive (Copyright Directive), as one of the topics still to be agreed on.

But when did all these start and how did copyright pop up in the AI Act discussion? The copyright consideration emerged following certain amendments introduced by the European Parliament back in the summer of 2023. The amendments required providers of foundation models to "*document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law*" [Article 28b(4c)]. Moreover, still in the European Parliament's position, the proposed Article 52(3) introduces certain transparency obligations surrounding AI-generated or altered content, requiring it to be labelled as such for the audience to be informed of its nature. However, in cases where the "*content forms part of an evidently creative, satirical, artistic or fictional cinematographic, video games visuals and analogous work or program*" such obligation will be limited to "*disclosing the existence of such generated or manipulated content in an appropriate clear and visible manner that does not hamper the display of the work and disclosing the applicable copyrights, where relevant*".

References to foundation models and copyright-related obligations were first integrated into the European Parliament's position on the file, contrasting with the absence of any similar reference in the Commission's initial proposal and the Council of the EU's position. The introduction of these provisions in the AI Act has sparked debates. On the one hand, organisations representing rightholders warmly welcomed the European Parliament's suggestions. The relevant stakeholders argue that such a disclosure of the works used by generative AI tools would allow them to regain control over their works. Through this disclosure, the rightholders will be able to decide whether they wish to continue, permitting the AI tools providers to use their works by concluding licences, or to opt-out by reserving their rights according to Article 4(3) of the Copyright Directive.

On the other hand, a view that such an obligation would be extremely challenging to implement, given the current state of the art, has been also expressed. The huge amount of datasets used to train the algorithms makes their disclosure especially burdensome. It has been further argued that the inclusion of fragmented copyright-related provisions in the AI Act, covering only certain aspects of copyright law, will not be an effective solution to addressing the issue. An overall revision of the current EU copyright framework seems rather more essential. As an upholder of this view, the Spanish Presidency of the Council, in a document shared with EU Member States' representatives in early September questioned "*whether AI Act should include provisions related to another legal framework that may need to be updated in the light of wide AI deployment*". Nevertheless, the Presidency's approach has been shifted since, in an effort to mediate between the EU Member States' representatives and the European Parliament's position. The latter has strictly defended its mandate, insisting that a non-inclusion of copyright transparency obligations would be unacceptable for them.

Following the latest trilogue on the AI Act, held in October, it appeared that a provisional agreement had been reached on the topic of foundation models and a tiered approach proposed by the European Parliament. Yet, in a subsequent technical meeting, France, Germany, and Italy called on the Spanish Presidency to withdraw its agreement. The Member State trio argued that regulation of foundation models would not be acceptable as it would hinder technological development. This twist was sufficient to put the negotiations on hold and make the European Commission intervene to propose a compromise. In the European Commission's compromise text, among the obligations for providers of general-purpose AI models, it suggests the latter "*put in place a policy to respect copyright law, in particular, to identify and respect the reservations of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 and draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model*".

This proposal marks the first official explicit reference to the need to reinforce the practical application of the TDM exception and ensure that rightholders' reservation of rights is fully respected. If adopted, rightholders' bargaining power will be significantly strengthened. The enforcement of the TDM exception effectively remains a challenge, due to the recent nature of the Directive and the great delay in its transposition by most Member States. What is certain is that we do not currently have clear evidence on how this provision can be effectively applied. Lack of targeted and harmonized standards and protocols leads us to believe that the provision may have already been rendered meaningless in practice. Therefore, the provision suggested by the European Commission serves as a reminder to the providers of generative AI tools to take the necessary measures to enforce their obligation under the Copyright Directive.

Text and Data Mining Exception – A brief memory refresh

Given that the TDM exception and the corresponding provision of Article 4 of the Copyright Directive have played a starring role in recent AI Act discussions, a closer look into its regime would be useful. Article 2(2) of the Copyright Directive defines text and data mining as "*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*". TDM is met in two provisions of the Directive, Articles 3-4 with the main focus being on Article 4, regulating TDM carried out for all purposes, while Article 3 regulates only the TDM carried out for scientific purposes.

Article 4 of the Copyright Directive reads:
"*Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.*

Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.

The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

This Article shall not affect the application of Article 3 of this Directive."

In the first paragraph, the provision provides for the rights which are subject to the exception. These include:

- temporary or permanent reproduction by any means and in any form, in whole or in part of a database (right provided to the database's author);

- extraction and/or re-utilisation of the whole or of a substantial part of a database (right provided to the database maker);

- reproduction right provided under the InfoSoc Directive for authors and various categories of rightholders with related rights;

- the permanent or temporary reproduction of a computer program by any means and in any form, in part or in whole;

- the translation, adaptation, arrangement, and any other alteration of a computer program and the reproduction of the results thereof, without prejudice to the rights of the person who alters the program;

- protection of press publications concerning online uses.

In short, the exception covers (1) reproductions and extractions of (2) lawfully accessible works (a criterion that needs to be interpreted flexibly including open access works, freely available content online, etc.), (3) for all purposes of TDM, (4) provided that the use of works has not been expressly reserved by their rightholders in an appropriate manner.

As noted, the most delicate part of the provision is the enforcement of the opt-out right, described in paragraph 3. Member States should have already put in place targeted measures to ensure that the rights reservation wouldn't be rendered meaningless in practice and that it would be carried out in a clear and precise way, to solve any uncertainty created for online service providers, including AI tool providers, seeking to identify how they can comply with their obligations.

Conclusions

The evolving landscape in the proposed AI Act and the discussion revolving around copyright aspects have spurred the pivotal question of whether the existing copyright regime is 'fit for the AI age' or whether it is incumbent for policymakers to examine a revision of the copyright framework in the upcoming mandate. Key parliamentarians, including MEP Dragos Tudorache (Renew, RO; AI Act co-rapporteur,) and Axel Voss (EPP, DE; AI Act shadow rapporteur and Copyright Directive rapporteur) have clearly stood in favour of such a revision, considering it necessary for the EU to keep pace with the technological developments in the field of AI.

In addition, the European Commission has acknowledged this challenge, signaling its readiness to look more closely into a potential revision. The creation of a database/repository collecting the opt-out statements to be made by rightholders would be a step in the right direction and would contribute substantively to the proper enforcement and respect of the opt-out rightholders's right. It is without doubt that the upcoming European Commission and European Parliament mandate springs surprises and their work both on AI and copyright enforcement will attract particular attention.