Warsaw, 28 February 2024

Andrzej Wyrobiec
Undersecretary of State
Ministry of Culture and National Heritage

Dear Mr. Minister,

We are submitting this position in response to the draft law on amendments to the law on copyright and related rights shared on 15th February 2024. We appreciate the opportunity to participate in this law's public consultations.

Our organization is a European think tank on the digital commons. We have been following policy debates on copyright and digital technologies at the European level and in Member States. Together with our partner organization COMMUNIA, the European Association for the Public Domain, we have been actively involved in the legislative debate on the Directive on copyright and related rights in the Digital Single Market and its implementation in the EU Member States.

In this submission, we want to focus on the articles related to implementing the text and data mining (TDM) exception.

On the rationale for excluding generative AI training from the scope of the TDM exception

The explanatory part of the implementation proposal notes, concerning the decision to exclude the use of works for training generative AI models from the scope of the TDM exceptions, that

> *The implementation of the directive now, in 2024, dictates that we refer here to the issue of artificial intelligence and whether text and data mining within the meaning of the directive also includes the possibility of reproducing works for machine learning. Undoubtedly, at the time the directive was adopted in 2019, the capabilities of artificial intelligence were not as recognizable as they are today, when "works" with an artistic and commercial value comparable to real works, i.e., man-made, are beginning to be created with the help of this technology. Thus, it seems fair to assume that this permitted use was not conceived for artificial intelligence. An explicit clarification is therefore introduced that the reproduction of works for text and data mining cannot be used to create generative models of artificial intelligence. [translation ours]*

We agree that the appearance of generative AI models as commercial products in the last two years has raised the need to clarify further the relationship between the EU copyright framework and the use of copyrighted works for training AI models. However, the available evidence does point in the opposite direction to the assumptions made in the justification quoted above. While there is little publicly available documentation of what lawmakers had in mind when they agreed on the structure of the TDM exceptions, what is available makes it clear that the development of artificial intelligence was explicitly factored into the discussions. Both the [European Parliament statement](#) and the [European Commission's explainer of the Directive](#), published after the adoption of the Directive in March 2019, specifically highlight that the TDM exception in Article 4 was introduced "in order to contribute to the development of data analytics and artificial intelligence."

In March 2023 (when the impact of Generative AI was more widely understood), the European Commission further clarified the relationship. In response to a [Parliamentary question](#)[1] that suggested that "The [CDSM] Directive does not address this particular matter," [Commissioner Breton pointed out](#)[2] that TDM exceptions do, in fact "provide balance between the protection of rightholders including artists and the facilitation of TDM, **including by AI developers**."

More recently, the [upcoming Artificial Intelligence Act](#)[3] contains a provision (in Article 52c(1)(c) that points out that developers of generative AI systems must "put in place a policy to respect Union copyright law, in particular, to identify and respect, including through state of the art technologies, the reservations of rights expressed pursuant to Article 4(3) of [the CDSM] Directive". In addition, the AI Act also contains a recital (Recital 60i) that explains the interaction between the training of generative AI systems and the exceptions in articles 3 & 4 of the Copyright Directive. This recital makes it clear that the EU legislators consider the use of works to train generative AI models to fall within the scope of these exceptions.

Compliance of the proposed implementation with the CDSM Directive

Excluding the reproductions made in the context of training generative AI models from the scope of the TDM exceptions proposed in Art. 26[2] and Art. 26[3] would result in a non-compliant implementation of the Directive.

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

---

[1] https://www.europarl.europa.eu/doceo/document/E-9-2023-000479_EN.html
[2] https://www.europarl.europa.eu/doceo/document/E-9-2023-000479-ASW_EN.html
[3]
https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf

not limited to, patterns, trends and correlations," the term Text and Data mining must be considered an autonomous concept of EU law that cannot be modified by the Member States in the legislation implementing the Directive. Doing so would undermine the very purpose of the Directive, which is to ensure the harmonization of exceptions and limitations across the member states. In this context, it must be stressed that should the current implementation proposals be adopted, Poland would be the only Member state with a modified scope of the text and data mining exceptions. There is a broad consensus that based on the definition contained in the Directive, the concept of text and data mining includes reproductions made to train generative AI models.

To avoid the Polish implementation being found to be non-compliant with the Directive, the following passages should be removed from the text of the proposed implementation act: "with the exception of the creation of generative models of artificial intelligence" (Art. 26[2]) and "except for the creation of generative artificial intelligence models" (Art. 26[3]).

A narrower TDM exception, if adopted in Polish law, will create an obstacle to developing generative AI solutions in Poland. The Polish policy on the development of artificial intelligence ("Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020") defines a range of goals that will be hindered by a narrow TDM exception. This includes increasing the supply of Polish AI solutions, strengthening Polish AI research, removing legal obstacles to AI development, and creating solutions dealing with specific challenges Poland faces (such as creating Polish language models). The last point is especially worth noting, as the challenges listed also concern issues that concern Polish culture and digital heritage - such as creating Polish language generative AI and researching its potential value and impact. In 2023, a consortium of Polish research institutions started working on PLLUM, an open-source, non-commercial language model. Further work on this model will become impossible if Poland adopts the TDM exception in the proposed, narrow form.

In this context, the proposal in Art. 26(2), to place out of scope of the TDM exception also non-commercial research conducted by research and heritage institutions is especially problematic. Research on generative AI is a key type of research into computational systems and digital technologies, which the Polish government should support.

There is an obvious need to balance various interests in the context of generative AI development: interests of entities building generative AI systems, interests of creators and other rights owners whose works are used in training models, and public interest. The approach proposed in the Directive and confirmed in the AI Act strikes this balance. The exception for text and data mining introduced in Art. 4 of the Directive offers a novel approach based on reservations of rights (*opt-outs*). This allows rights holders who want to reserve rights to do so – and potentially also offer licensing options for their works – while enabling TDM research and AI training on a broad pool of publicly available works.

Compatibility with the copyright provisions in the upcoming Artificial Intelligence Act

As highlighted above, the upcoming AI Act, which will directly apply in Poland, contains several provisions to ensure compliance of providers of general-purpose AI models (including generative AI models) with the EU copyright framework . These are based on the understanding that using copyright-protected works for training such models falls within the scope of the exceptions in articles 3 and 4 for the copyright directive. Having a Polish implementation of the Directive that directly contradicts assumptions underpinning the relevant portions of the AI Act will result in considerable legal uncertainty.

Challenges related to the practical applications of Article 4(3) of the CDSM Directive

The TDM exceptions in the CDSM Directive have been drafted to ensure a carefully crafted balance between the interests of rightholders and the interest of society at large to engage with an increasing amount of public information and culture through means of computational analysis (including but not limited to the current wave of AI technologies). This balance is assembled from four elements:

1. Both TDM exceptions only apply to lawfully accessible works;
2. TDM for the purpose of scientific research is otherwise allowed without restrictions, ensuring a privileged position for public knowledge institutions;
3. Rightsholders can reserve their right to allow TDM for any purpose other than scientific research, allowing them to control if and how their works can be used;
4. All other works can be used from TDM by anyone for any purpose.

This system ensures that works and other copyrighted matters not actively managed by their rightsholders remain available for TDM uses. This opt-out approach has been explicitly designed to prevent an orphan works-type situation in which information and cultural expression remain off limits for further use because rightsholders have been assigned exclusive rights they do not intend to exercise. This mechanism addresses one of the key shortcomings of the existing copyright framework and is also present in other parts of Directive[4].

There is currently a lot of uncertainty with regard to how the machine-readable rights reservation in line with Article 4(3) of the Directive can work in practice. While various AI developers have started offering opt-outs from training data collection, there are currently no

---

[4] Compare Article 8(4), which allows rightsholders to opt out from the mechanism allowing the use of Out-of-commerce works and other subject matter by cultural heritage institutions.

generally recognized standards or protocols for the machine-readable expression of the reservation[5].

Here, there is an urgent need for the European Union and its Member States to provide a standardized framework for making such opt-outs more accessible, efficient, and scalable both for rightsholders and AI companies. Such an intervention, in combination with the obligations introduced in the AI Act, would offer those rightsholders who want to exclude or license their works from use for training generative AI models (or license them for this purpose) the same legal protections as a blanket exclusion of generative AI training from the scope of the TDM exceptions. While ensuring access for scientific research and access to works where creators and other rightsholders have no objections against such uses.

How to strengthen the position of creators and rightsholders within the framework provided by the CDSM directive

In general, the system introduced by the CDSM TDM exceptions allows creators to require remuneration for using their works for AI training outside of academic research contexts. This requires them to opt out and offer licenses to use their works. Given the scale of AI training and the amount of work involved, such licenses would need to be offered as collective licenses by organizations representing creators or by large institutional rightsholders (such as publishers) with large portfolios. We are already seeing collective management organizations across the EU opting out on behalf of their members to license their works for AI training uses[6]. Such forms of voluntary collective management, on top of the opt out framework, are a suitable mechanism to address the economic interests of organized rightsholders and should be supported by creating the conditions for transparent and efficient distribution of licensing revenues.

One novel aspect of the current generation of generative AI models is the fact that they are trained on a very large number of publicly accessible works. This raises questions related to the sustainability of public information resources that are used for such training and for which their creators and or rightsholders do not see a need to opt out. In order to account for this, a charge on revenues of providers of generative AI models trained on publicly available data should be introduced. Such a financial mechanism would guarantee that entities benefitting from shared and publicly accessible resources contribute to the

---

[5] For more detail see our [Policy brief on defining best practices for opting out of ML training](#) from September 2023.
[6] See for example the Dutch visual Arts CMO Pictoright:
https://pictoright.nl/nieuws/collectieve-opt-out-pictoright-aangeslotenen/ or SACEM in France:
https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out

maintenance and future creation of these resources. The revenues obtained through such a charge should feed into a public fund supporting a broad range of initiatives supporting the cultural and knowledge commons. This could include individual artists and creatives, cultural organizations, public institutions stewarding the digital commons (for example, digital libraries or open-access journal repositories), or civic initiatives (like Wikipedia). Such a fund would preferably be European in scope, but a system of national funds is also possible. The Polish Film Institute's (PISF) support for Polish cinematography, funded with payments collected from entities in the broadly understood film sector, is a good reference point for this approach. The German Sovereign Tech Fund is another example of a public fund aimed at supporting the digital commons. It should be noted that this type of fund is not intended as a form of remuneration for individual creatives and rights holders for the use of their work for training generative AI (this objective would be achieved through the opt-out mechanism as described above). Instead, it is meant to provide economic support to the cultural and knowledge commons, understood as a collective good.