EARE's Position on the Own-initiative Report on Copyright and Generative AI

The European Alliance for Research Excellence (EARE) would like to share its views on the own-initiative report on copyright and generative AI authored by MEP Axel Voss for the JURI Committee. As an organization representing researchers and innovators in the EU, this paper aims to point out the impact that some elements of the INI report can have on the research and innovation ecosystem, especially when research and innovation are essential to drive economic growth, as reflected by [Mario Draghi](#) and [Enrico Letta](#)'s report.

For researchers, universities, start-ups and innovators, open data policies and the Text and Data Mining (TDM) exceptions within the Copyright Directive are key to expand knowledge and innovation in Europe, and provide an environment that leads to a skilled AI workforce. It is our view that the European Union must ensure that the current copyright framework fits the needs of all stakeholders. Suggesting that the current TDM exceptions require a reinterpretation is short sighted and imposes additional barriers and costs for research and innovators.

Open data policies are key for research, innovators and startups

Open data policies and the Text and Data Mining (TDM) exceptions in Articles 3 and 4 of the DSM Directive are essential for researchers, innovators and startups to ensure research excellence, avoid bias in research, and develop new technologies such as AI.

TDM exceptions apply to AI systems, including generative AI

The report claims that the training of generative AI systems is not currently covered by the existing TDM exception. This challenges the interpretation provided by the AI Act and by multiple statements by the Commission and should therefore be avoided.

TDM exceptions were intended to train AI systems, including generative AI. The opt-outs within Article 4 of the DSM Directive were expressly intended to allow rightsholders to reserve their rights on TDM activities for commercial purposes.

The DSM Directive defines TDM as "*any automated analytical technique aimed at analysing text and data in digital form to generate information*". This already covers training AI models, including those used for generative AI, which involves analysing large datasets to extract patterns.

The application of the TDM exception to train AI systems is not a new issue for policymakers and EU institutions and Member States have been aware of these discussions for long time. This is reflected in the negotiations of the DSM Directive and the AI Act. Similarly, Articles 3

and 4 of the DSM Directive as well as the Recital 105 of the AI Act explicitly confirm that TDM exceptions apply when training general purpose AI models.

Today, researchers often refrain from using research tools due to fears of copyright infringement and the fragmented implementation of the Copyright Directive across Member States. The new interpretation established in the report adds to this uncertainty, creating further confusion for researchers and innovators when using data.

Rather than new laws, the focus should be on properly applying the current framework

The introduction of a dedicated exception, distinct from that provided for TDM under Article 4 of the DSM Directive, or expanding the scope of the provision to encompass the training of genAI as proposed in the INI report (paragraph 7), would add **unnecessary complexity and confusion for researchers and innovators**, further complicating an already complex regulatory framework. This would prevent researchers and innovators from using content they already have access to, renegotiating access to content for AI purposes, reducing the data available, increasing bias in research, raising costs, and undermining the willingness among researchers and innovators to use data. Ultimately, this would negatively impact the quality of scientific research in the EU.

As mentioned previously, the current definition of TDM is broad enough to include generative AI training and the EU AI Act (Recital 105), as well as several statements from the European Commission confirm this interpretation.

Instead of new legislation, EU policymakers should prioritize the proper implementation of TDM exceptions and the Copyright Directive while guaranteeing that the current framework respects rightsholders' reservation of rights. The key focus should be on facilitating and improving the implementation of the current legislation across Member States instead of creating additional norms which will add more uncertainty, particularly for researchers, innovators, and startups.

New rules or legislative changes will add further complexity to the copyright legislation in the EU, particularly undermining the work of researchers, innovators, and startups.

The European Commission should refrain from introducing any measures before the comprehensive review of the Copyright framework scheduled for 2026, as this will add further confusion and potentially distort the review. Future efforts to clarify the implementation of the current copyright framework should come through non-binding instruments such as guidelines, which should consider the specific needs and realities of the research and innovation ecosystem.

EUIPO Central Register of Opt-outs

EARE recognizes the potential efficiencies with a centralized EU-level register of opt-outs as a tool to help researchers and innovators identify which data can be accessed for commercial research, as well as AI model development and improvement**. It is paramount that any registry is consistent with existing provisions of the EU DSM**, including that a registry satisfies the requirements of Article 4 of the EU DSM that a reservation of rights be expressed in an appropriate manner, such as a machine-readable format. This includes the integration with other systems, including research tools, and AI training pipelines, and an Application Programming Interface (API) which will allow researchers and developers to receive updates. It is important that such a register does not undermine the unconditional TDM exception that exists under Article 3 or overlook the existence of other relevant exceptions.

However, a centralized system within the EUIPO could become a barrier to the diversity and richness of data for researchers and innovators, especially if it creates compliance burdens. These burdens should be carefully assessed and mitigated. Single researchers may face difficulties to access this register and to understand if opt-outs apply to their research projects in the context of public-private partnerships. Similarly, researchers, SMEs, and startups often lack legal and technical resources needed to follow and monitor the updates of the register. To facilitate the work of researchers and innovators, the register should also be designed for ease of use, with particular attention to the needs of researchers, startups, and SMEs. The managing entity should be equipped to handle multiple calls and users at the same time, and provide continuous assistance, particularly to researchers, startups, and SMEs. Moreover, the opt-outs must be regularly updated to reflect changing preferences. The obligation to continuously monitor the register or implement measures to comply with updated register opt-outs can impose a burden for researchers and startups. Most importantly, the register must offer robust assurances that opt-outs accurately reflect the rights of all owners of a particular work. This is essential to prevent future legal disputes that could hinder research and innovation.

EARE also acknowledges the idea of involving EUIPO as a trusted intermediary. However, researchers and startups might face increased bureaucracy and compliance requirements, when using datasets for training AI models. This could slow down research timelines. As the European Commission has embraced the reduction of administrative barriers and EU's competitiveness, the involvement of the EUIPO as a trusted intermediary should be rightly evaluated.

Transparency

Transparency and reproducibility of research is core to the values of EARE. The INI report considers on recitals O to Q that full transparency should consist "*in an itemised list identifying each copyright-protected content used for training*". This approach is not proportionate or practical. Instead of new provisions on transparency, **the EU should focus on the right implementation of the current AI Act**, the Code of Practice and the template of the summary of training data. Before introducing any additional requirements on transparency, these tools should be rightly implemented.

Obligations which touch upon transparency and openness of AI tools need to be **proportionate**, and any measures should be properly understood, and balanced. It is important to note that the European Commission's study on improving access to reuse of research results for scientific purposes, mentioned that a detailed summary of the data used for training can "*add a layer of compliance costs for research organizations*". Similarly, for SMEs, and startups, this document can be difficult to implement. For this reason, the European Commission should focus on working with AI providers, research organizations, SMEs, startups, and other relevant stakeholders to monitor the implementation of the template and improve it if needed before the entry into application of the enforcement powers of the AI Office on 2 August 2026. Similarly, it is essential for SMEs and startups the **protection of trade secrets** to secure funding. When working with the AI Office to assess the submissions from SMEs and startups, the AI Office should ensure that sensitive or proprietary information is not disclosed.

Regarding the **irrebuttable presumption** introduced in the INI report, which assumes that copyrighted works have been used in training AI models if transparency requirements are not fully met, EARE considers that this could **negatively impact access to data for researchers and innovators**. This approach risks disproportionately affecting researchers and start-ups located in the EU, particularly those lacking the resources to ensure full compliance or who may be unaware of these obligations despite being willing to comply. Since the presumption is irrebuttable, it does not allow small developers or researchers to prove otherwise, even if they have not included opted out works to train AI models. This presumption could further deter researchers and smaller AI startups to develop AI models. Similarly, since the irrebuttable presumption would apply to any form of TDM, it could impact the work of researchers and innovators by discouraging the use of data and potentially facilitating bias into research results. As previously mentioned, the EU should focus on the right implementation of the current provisions on transparency established in the AI Act, and the future Code of practice.

Remuneration Conditions

The report calls on the Commission to immediately impose a remuneration obligation on general-purpose AI models and systems for the use of content protected by copyright (Paragraph 4), with "*such obligation applying until the reforms envisaged in the report are enacted*".

However, this provision risks undermining data access for researchers, cultural institutions and innovators. The TDM exception clarifies that the use of a copyrighted work to train an AI model is not a copyright infringement and therefore does not require remuneration. Further, immediate remuneration obligations would make it harder or more expensive for researchers and innovators to access large datasets, putting ongoing research projects at risk, and compromising future scientific research and innovation.

Similarly, many research projects involve public-private partnerships. If the provision is not clear enough, these partnerships might be treated as commercial, even if their purpose is for non-commercial scientific research. This could also prevent researchers and innovators to be involved in public-private partnerships.

This immediate obligation may also distort the future review of the copyright framework expected in 2026.

About EARE*: The European Alliance for Research Excellence (EARE) was convened by Microsoft in 2017, and now brings together nine members from the research and innovation ecosystem in Europe, including the Association of European Research Libraries (LIBER Europe), the European Bureau of Library, Information and Documentation Associations (EBLIDA), BSA | The Software Alliance, Microsoft, Allied for Startups, LACA, Research Libraries UK, SCONUL (Society of College, National and University Libraries), and UCL (University College London) Library, advocating for the EU to live up to its innovation potential in the digital economy.*