

EAIF SUBMISSION

JUNE 2026

Response to the European Commission's Call for Evidence

Review of Directive (EU) 2019/790 on Copyright in the Digital Single Market (CDSM Directive)

SUBMITTED BY

European AI Forum (EAIF)

MEMBERS REPRESENTED

14 national AI associations ·
3,000+ members

SCOPE

AI training & TDM provisions

Executive Summary

The European AI Forum (EAIF), representing 14 national AI associations and more than 3,000 members across the EU, submits this response to the Commission's Call for Evidence on the review of the CDSM Directive. We welcome this timely review, which arrives at a critical juncture: the deployment of General Purpose AI (GPAI) models has exposed significant interpretive gaps in Articles 3 and 4 CDSM, creating legal uncertainty that disproportionately burdens European AI developers relative to their global competitors.

Our key positions are: (i) Articles 3 and 4 CDSM encompass the full AI training pipeline, including data acquisition, preprocessing, and iterative training; (ii) Article 3's research exception permits the open-source or public release of trained models; (iii) Article 4 is, in spirit, a fully harmonising provision and national deviations run counter to its objectives; (iv) the mismatch between the AI Act's copyright compliance assumption and the CDSM's practical framework must be resolved legislatively; (v) opt-out protocols require Commission-level standardisation; and (vi) proportionality for SMEs must be structurally embedded. The EAIF calls for targeted legislative clarification rather than a wholesale recast.

Section 1: About the European AI Forum and This Submission

The European AI Forum (EAIF) is the umbrella organisation of 14 national European AI associations and clusters, collectively representing more than 3,000 members — including AI startups, scaleups, established enterprises, research institutions, and individual experts active in artificial intelligence across the European Union.

As the largest European AI stakeholder organisation, the EAIF engages systematically with EU legislative and regulatory processes affecting the development, deployment, and governance of AI systems. Our membership spans the full AI value chain: model developers, application providers, infrastructure operators, and downstream deployers. The breadth of our membership positions us to offer a cross-sectoral perspective on the intersection of intellectual property law and AI innovation.

This response focuses specifically on the application of the CDSM Directive to AI training activities. It reflects the operational realities faced by our members when building and deploying AI systems in Europe, and the legal uncertainty they currently navigate. We do not comment on aspects of the Directive unrelated to AI.

The EAIF submits this response in the spirit of constructive regulatory cooperation. We are committed to an IP framework that is both rights-respecting and innovation-enabling — objectives that, properly designed, are complementary rather than in tension.

Section 2: Articles 3 and 4 CDSM — Legal Interpretation for AI Training

2.1 Scope: Both Articles Cover the Full AI Training Pipeline

A foundational question for this review is whether Articles 3 and 4 CDSM apply only to the act of ingesting training data, or to the full AI training pipeline. The EAIF's position is unequivocal: both provisions must be interpreted as covering the entire pipeline, including:

1. **Data acquisition and crawling** — the initial retrieval and storage of content for training purposes;
2. **Data preparation and preprocessing** — filtering, cleaning, tokenisation, vectorisation, and the creation of intermediate datasets;
3. **Model training** — the iterative optimisation process during which model weights are adjusted based on the training corpus;
4. **Fine-tuning and continued pre-training** — subsequent training runs using additional or curated datasets.

This interpretation follows from both the text and the purpose of the Directive. Recital 8 CDSM explicitly states that text and data mining "requires the reading, copying, extraction, and analysis of information in

large amounts." Each step in the AI training pipeline satisfies at least one of these criteria. An interpretation that exempts data ingestion while leaving preprocessing and training exposed would be commercially irrational and technologically illiterate: the reproduction involved in storing a tokenised dataset is legally indistinguishable from the reproduction involved in storing the raw corpus.

The Commission should clarify this in the review outcome, whether through explanatory guidance or, if legally necessary, through a targeted amendment. Legal uncertainty at this foundational level creates chilling effects, particularly for smaller developers who cannot absorb litigation risk.

2.2 Article 3 Research Exception: Public Release of Trained Models Is Permitted

Article 3 CDSM provides a mandatory exception for text and data mining carried out by research organisations and cultural heritage institutions for the purposes of scientific research. A question of direct commercial significance is whether a model trained under the Article 3 exception may subsequently be released publicly or as open source, or whether such release retroactively vitiates the exception.

The EAIIF's position is that Article 3 permits the public or open-source release of models trained under the research exception, provided the following conditions are met:

- The training activity itself was conducted for scientific research purposes;
- The stored copies of training data are not distributed independently;
- The released artefact is the trained model (its weights and architecture), not the underlying training corpus.

This interpretation is consistent with the structure of the exception, which conditions the benefit on the purpose of the mining activity (scientific research), not on the subsequent fate of any downstream artefact that does not itself constitute a reproduction of the original works. A trained model is a statistical transformation, not a copy. To hold otherwise would render the research exception commercially sterile and would deter precisely the kind of open, reproducible AI research that the EU's scientific community and AI ecosystem depend on.

The Commission should issue interpretive guidance confirming this position. Failure to do so will push open-source AI research outside the EU or force researchers into commercially unsustainable licensing arrangements.

2.3 Article 4 and Harmonisation: National Deviations Run Counter to the Directive's Objectives

Article 4 CDSM provides an exception for text and data mining for purposes other than scientific research, subject to the rights-holder's right to opt out. Unlike Article 3, Article 4 is available to any person, including commercial AI developers, and is subject to the reservation that rights holders may restrict TDM for commercial purposes.

A critical interpretive issue is whether Article 4 is fully harmonising, i.e., whether Member States may modify its scope — expanding or restricting coverage — through national law. The EAIIF's position is that

Article 4 is, in spirit, a fully harmonising provision and that national deviations run counter to its objectives.

This follows from the core logic of EU copyright harmonisation: where a Directive provides a uniform exception with specific conditions, additional restrictions or alternative regimes layered on top by Member States undermine the Single Market objective. The so-called "Polish model" — under which Poland has sought to modify the conditions for invoking Article 4 through national implementing legislation — represents a significant departure from the harmonised standard that the Commission should address.

Single Market concern: A fragmented Article 4 landscape, in which the commercial TDM exception has different practical content in different Member States, directly impedes the Single Market objective the CDSM Directive was enacted to serve. AI developers building training datasets from multilingual, cross-border corpora cannot operationalise an exception that varies in scope across 27 jurisdictions.

The Commission should:

- Confirm that Article 4 should be applied in its harmonising spirit across the EU and call on Member States to align national implementations accordingly;
- Initiate a structured dialogue with Member States whose national implementing laws deviate from the harmonised standard;
- Consider whether a Regulation is the appropriate instrument for any revised TDM framework, precisely to prevent this category of fragmentation.

Section 3: Coherence Between CDSM and the AI Act (Art. 53 GPAI + Recital 105)

The AI Act, which entered into application in stages from August 2024, imposes on providers of GPAI models an obligation under Article 53(1)(c) to "put in place a policy to comply with Union copyright law." Recital 105 further elaborates that this requires GPAI providers to identify and respect opt-outs "in accordance with Article 4(3) of Directive (EU) 2019/790."

This creates a structural coherence problem. The AI Act assumes a functioning, operationally clear CDSM Article 4 regime as its underlying copyright compliance layer. The CDSM, as currently interpreted, does not deliver that clarity.

The AI Act's three unmet assumptions

- The AI Act assumes that opt-out protocols are standardised and machine-readable — they are not.
- The AI Act assumes that the scope of Article 4 is uniform across the EU — it is not.
- The AI Act assumes that the training pipeline activities covered by Article 4 are settled law — they are not.

The consequence is that GPAI providers — particularly those without large legal departments — cannot operationalise the Article 53 AI Act obligation in a technically and legally reliable manner. This is not a hypothetical concern: it is the daily operational reality of our members.

The gap between the AI Act's assumption and the CDSM's practical framework must be closed legislatively. The Commission cannot resolve this through soft guidance alone. We recommend that the CDSM review include a specific provision or recital that directly references the GPAI context, confirming the application of Articles 3 and 4 to AI training activities and providing a normative bridge to Article 53 AI Act compliance.

Section 4: Opt-Out Protocol Standardisation

Article 4(3) CDSM grants rights holders the right to opt out of the commercial TDM exception "in an appropriate manner." This formulation — deliberately technology-neutral at the time of drafting — has become a source of significant legal and operational uncertainty in the AI context.

Multiple competing opt-out protocols now exist in the market:

- **TDMRep** (developed under W3C auspices) — the most technically rigorous, machine-readable, scalable protocol;

- **robots.txt** — widely deployed but designed for search engine crawling, not copyright-aware TDM;
- **TDM-AI** — a more recent entrant with partial industry adoption.

The proliferation of protocols creates a situation in which AI developers cannot determine with legal certainty whether a given data source is subject to a valid opt-out; rights holders cannot be confident that their expressed opt-out will be honoured by all crawlers; and litigation risk is distributed asymmetrically.

The Commission must take the following actions:

- Publish an official list of Commission-recognised opt-out protocols. TDMRep should be designated as the primary recognised protocol, given its technical maturity and W3C backing;
- Impose a multi-standard interoperability requirement: AI developers must check for opt-outs expressed via any Commission-recognised protocol;
- Establish a transitional recognition period for legacy opt-outs expressed in non-standardised formats;
- Consider delegating technical maintenance of the recognised protocol list to the EUIPO or a designated standard-setting body.

Section 5: Constructive Licensing and Fair Remuneration

The EAIF recognises that the legitimate interests of rights holders must be addressed structurally, not merely through opt-out mechanisms. A well-functioning licensing framework is in the long-term interest of both AI developers and the creative sector.

We support the exploration of extended collective licensing (ECL) mechanisms under Article 12 CDSM as a complementary pathway for AI training use cases, particularly where individual rights clearance is operationally impractical at scale. ECL arrangements, where functioning collecting societies exist and representative coverage can be demonstrated, offer a rights-respecting alternative that also generates remuneration flows to creators.

We also support a prospective exclusion model: rights holders who have not opted out prior to a training run — and whose works are incorporated into a training corpus in good faith — should not be able to assert retroactive infringement claims. Prospective exclusion, combined with fair compensation via ECL, provides a commercially workable framework that balances innovation and rights protection.

For SMEs, in particular, licensing frameworks must be proportionate: collective licensing bodies should be required to offer standardised, tiered tariff structures accessible to small developers without requiring bespoke negotiation.

Section 6: SME Proportionality

The EAIF's membership is predominantly composed of SMEs and startups. We note with concern that the compliance burden associated with the current CDSM TDM framework — to the extent it is operationally defined at all — is structurally calibrated to organisations with dedicated legal and data engineering teams.

Compliance with opt-out protocols requires: automated crawling infrastructure capable of checking multiple protocol formats; legal review of each dataset's provenance and licensing status; ongoing monitoring as rights-holder opt-out decisions change over time; and documentation systems sufficient to demonstrate good-faith compliance in any subsequent litigation or regulatory inquiry.

The Commission must embed structural proportionality into any revised CDSM TDM framework. Specifically, we recommend:

- **A safe harbour for good-faith SME compliance:** SMEs that implement recognised opt-out protocol checks and maintain reasonable documentation of their data sourcing practices should benefit from a presumption of compliance, reversible only by evidence of wilful infringement;
- **Harmonised documentation templates** (analogous to the AI Act's GPAI code of practice model) that SMEs can use without bespoke legal advice;

- A **Commission-maintained public registry of pre-cleared open datasets** suitable for AI training, reducing the due diligence burden for SMEs that rely on common training corpora.

Section 7: The Three-Step Test and AI Training

Article 5(5) of Directive 2001/29/EC (InfoSoc), incorporated by reference into the CDSM Directive framework, conditions copyright exceptions on the "three-step test": exceptions apply only (i) in certain special cases, (ii) that do not conflict with a normal exploitation of the work, and (iii) that do not unreasonably prejudice the legitimate interests of the rights holder.

There is a risk — already visible in litigation strategy by some rights-holder interests — that the three-step test will be used as a ground for denying the application of Articles 3 and 4 CDSM to AI training activities. The EAIF cautions against this approach.

The EAIF's position on the three-step test in the AI training context:

- **AI training is a "special case":** It is a specific, technically defined activity with a clear and bounded purpose — the optimisation of model weights. It is not general-purpose consumption of copyrighted works.
- **AI training does not conflict with the normal exploitation of the work:** The "normal exploitation" of a novel, a news article, or a scientific paper is reading, licensing for reproduction, or adaptation. AI training is a qualitatively different use: the model does not reproduce the work, it learns statistical patterns from it.
- **Generative AI outputs are not market substitutes for training works:** A GPAI model trained on literary works does not "substitute" those works in any commercially meaningful sense. A language model's output is a probabilistic generation conditioned on billions of parameters — not a copy of any training document.

The Commission should confirm in the review outcome that the three-step test — applied consistently with the Berne Convention — does not serve as a blanket override of the Articles 3 and 4 exceptions in the AI training context, and that harm must be demonstrated on the specific facts rather than presumed from the category of use.

Conclusion and Recommendations

The EAIF supports a CDSM framework that provides genuine legal certainty for AI developers while respecting the legitimate interests of rights holders. These objectives are compatible, but they require the Commission to act with clarity and specificity.

The EAIF calls on the Commission to adopt the following recommendations as part of the CDSM Directive review:

- 1 Confirm that Articles 3 and 4 CDSM apply to the full AI training pipeline, including data acquisition, preprocessing, and iterative training, through binding guidance or targeted amendment.
- 2 Clarify that Article 3 permits the public and open-source release of models trained under the research exception, provided the training data itself is not independently distributed.
- 3 Confirm that Article 4 should be applied in its harmonising spirit across the EU and call on Member States to align national implementations accordingly.
- 4 Publish a Commission-recognised list of opt-out protocols, designating TDMRep as the primary standard, and impose a multi-protocol interoperability obligation on AI developers.
- 5 Adopt a legislative bridge between CDSM Article 4 and AI Act Article 53 to close the coherence gap and provide GPAI providers with an operationally workable copyright compliance framework.
- 6 Explore extended collective licensing (ECL) under Article 12 CDSM as a complementary licensing pathway for AI training, with a prospective exclusion model and proportionate SME tariff structures.
- 7 Introduce structural SME proportionality through a good-faith compliance safe harbour, harmonised documentation templates, and a public registry of pre-cleared training datasets.
- 8 Confirm that the three-step test — applied consistently with the Berne Convention — does not presumptively override Articles 3 and 4 in AI training contexts, and that market substitution must be demonstrated on specific facts rather than inferred from the category of use.
- 9 In any future reform of the TDM framework, give serious consideration to whether a Regulation — rather than a revised Directive — would better prevent fragmentation, while recognising that the primary objective should be targeted clarification of the existing framework rather than a wholesale recast.

Submitted on behalf of the European AI Forum (EAIF)

Brussels, 24 June 2026

Contact: info@eaiforum.org

The European AI Forum (EAIF) is the umbrella organisation of 14 national European AI associations and clusters, representing over 3,000 members including SMEs, corporates, research organisations, and individual experts in artificial intelligence. EAIF engages systematically with EU legislative and regulatory processes on behalf of the European AI ecosystem.