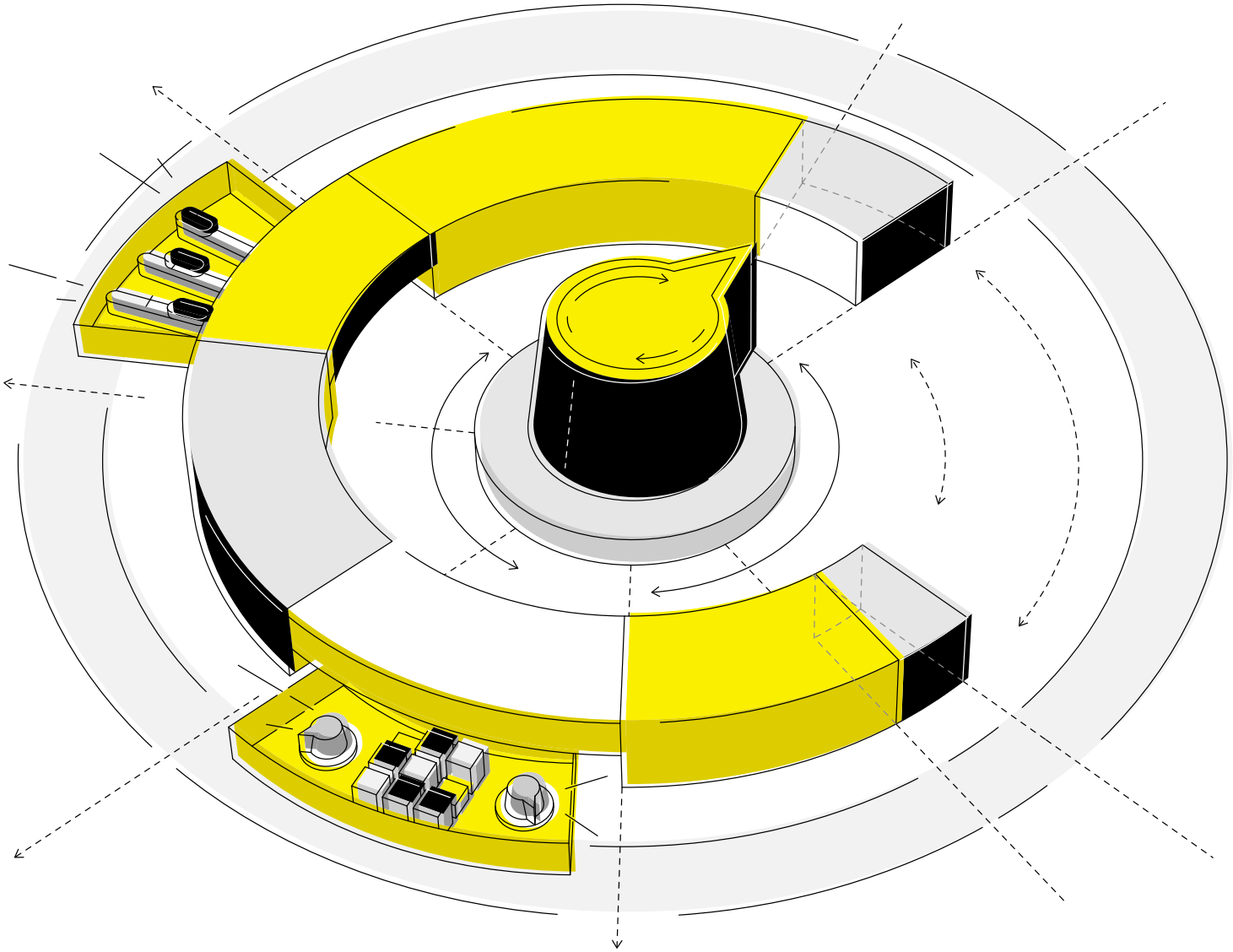
SHARING THE VALUE OF AI TRAINING

Proposal for a collective remuneration mechanism



OPEN _FUTURE

NOVEMBER 2026

Paul Keller

In June 2025 Open Future published Beyond AI and Copyright, which argued that generative AI poses structural risks to the information ecosystem that copyright-based remedies cannot address on their own. It proposed a levy on commercial AI services to fund the people and institutions that produce and steward public knowledge.

The Commission's targeted consultation on a better copyright environment for European creativity and innovation asks how EU copyright law should respond to the use of protected content by generative AI. This paper is written as a contribution to that consultation. It builds on Beyond AI and Copyright but is narrower: it proposes a mechanism tied to the training of AI models, built with the instruments of EU copyright and market-access law.

The paper does not propose legislative text. It sets out the requirements an intervention has to meet, then breaks the mechanism into six constituent parts: what triggers the obligation, how it operates, what it is calculated on, how the proceeds are distributed, how it relates to existing copyright rules, and what legal form it takes. For each part we state what we propose and identify any questions that remain to be settled. These are genuinely open questions that can be answered, so they should not be read as reasons not to proceed.

CONSIDERATIONS

In the call for evidence that opened this process in May 2026, the Commission states that it is exploring ways to "enhance the licensing and enforcement of copyright and related rights in the AI context, improve the conditions for creators' remuneration, while making it easier for providers of generative AI to access copyright-protected content", taking into account the impact on all affected parties, including start-ups and SMEs. This paper treats these objectives as requirements that an intervention has to meet, and starts from a broader one that frames them: sustaining the information ecosystem on which AI models are trained.

A sustainable and diverse information ecosystem

Generative AI models are trained on the accumulated output of the open web and of digitised collections. That includes works managed by rightholders, but also the public domain, openly licensed material, publicly funded research, government information, and decades of everyday expression by people who never expected payment. Parts of this content, such as the public domain and openly licensed material, form a commons. Much of it is protected by copyright and has been made publicly available by its creators and rightholders. The continued production of all of this content depends on creators, publishers, archives, libraries, public broadcasters and community projects, all of which are now under pressure from the AI systems trained on their work.

A licensing market, no matter how well it functions, cannot reach most of this content. Licensing works where a rightholder can be identified, has the capacity to negotiate, and holds rights in the material concerned. For the bulk of what models are trained on, none of these conditions is met. An intervention that aims to sustain the information ecosystem therefore cannot rely on

licensing alone. It needs a new collective mechanism, built from instruments that copyright law already uses — collective management and statutory remuneration — to do what individual licensing cannot.

Fair remuneration for creators and rightholders

An intervention that improves the conditions for creators' remuneration has to reach creators and rightholders whose works are part of the publicly available content on which commercial AI models are trained. Training on these works is essential to model development: training shapes the models' capabilities and behaviours, and the resulting value is captured by the companies that supply those models commercially. The issue becomes more pressing if AI systems compete with the creators whose works they were trained on, eroding the labour markets that have funded creative and journalistic work. Yet under the current framework creators and rightholders are not sharing in that value. The opt-out in Article 4 of the CDSM Directive was meant to give rightholders leverage to license. In practice, organisations representing creators report that exercising a reservation produces neither a licence nor a payment — only an unverifiable expectation that the reservation will be respected. Enforcing it is left to individual creators and small rightholders, who are least able to carry that burden.

A system that works beyond the largest players

Where licensing does happen, it happens between the largest AI developers and the largest rightholders. Those deals are the easy part. They leave out the long tail of individual creators and small rightholders, who lack the bargaining power to license. They also leave out smaller AI developers. Smaller European developers are squeezed from both sides: they cannot match the prices the largest developers pay for licences, and they cannot afford the legal risk of training without them. A workable mechanism has to function at both ends of the market, not only for the parties that can already negotiate.

Continued access to training data

European AI developers need access to the same breadth of training data as their competitors elsewhere. A framework that makes access conditional on licensing every work, or on navigating millions of individual reservations, would put that access out of reach for all but the largest developers, most of which are not European. An intervention therefore has to guarantee all players access to all of the publicly available information for training on terms that ensure it is paid for.

This matters most for open-source development. Open models can be run on European infrastructure, inspected and adapted to European languages and public tasks, and much of the European AI ecosystem, from start-ups to publicly funded projects such as OpenEuroLLM, builds on them. Yet their developers face the current legal uncertainty in its sharpest form. The AI Act's exemptions for open-source models do not extend to the copyright obligations. The research exception does not cover most of the companies, foundations and communities that develop

open models. And because these developers publish their models and often document their training data, their practices are the easiest to challenge.

Legal certainty and a level playing field

European AI developers face legal uncertainty about how reservations must be expressed, identified and respected at scale, and about how EU rules apply to models trained outside the EU. Honouring reservations work by work is costly and unreliable at the scale at which models are trained: reservations are often not expressed in a form that can be read automatically, they often do not identify specific works, and the same work circulates in many copies, including on sites the rightholder does not control. These costs fall hardest on smaller providers. Any intervention has to provide clear rules that apply equally to every provider supplying models on the European market, wherever the model was trained.

Reconciling the competing objectives

These objectives pull in different directions. Making training on publicly available content legally certain and broadly accessible, and ensuring that creators and rightholders are actually paid, are both goals the Commission has set for this initiative, and they are in tension. Measures aimed at one tend to come at the cost of the other: stricter reservation regimes make access harder, while broader exceptions leave rightholders without remuneration. Adding layers of each increases complexity without resolving the conflict.

The way out is to stop trying to reconcile these objectives transaction by transaction, and to do it collectively instead. A remuneration obligation attached to the commercial supply of AI models on the EU market, paid into a fund and distributed to creators, rightholders and the stewards of publicly available content, allows training on publicly available content to proceed with legal certainty while ensuring that it is paid for. In exchange, training on publicly available content is no longer subject to individual reservations. The rest of this paper sets out how such a mechanism would work.

1. TRIGGER AND SCOPE

The proposed obligation is triggered by the commercial supply, on the EU market, of access to AI models trained on publicly available content. The obligation applies to whoever first commercially supplies access to these models in the EU, wherever that company is established and wherever the model was trained.

The obligation attaches to the commercial supply of the model in the EU, not to the training. It does not depend on where training took place, under which law it was lawful, or whether a reservation was expressed. What is remunerated is the commercial supply on the European market of a model built on publicly available content. Commercial supply means making the model available, or using it in products or services, in the course of a commercial activity, whether in return for payment or free of charge. It covers providers that use their own models in

their own products as well as third parties that host or integrate open models. That is something the EU can regulate without reaching training or other development acts that took place elsewhere.

Publicly available content means content that anyone can access without creating an account, entering a contract, paying, or circumventing a technical protection measure. It is a subset of the content to which providers have lawful access under Article 4 of the CDSM Directive. Content behind subscriptions, paywalls or technical protection stays outside the mechanism and remains available for licensing. This includes content that can be viewed openly but can only be obtained by circumventing an effective technical protection measure, such as video delivered through protected streams. Terms of service that restrict what may be done with openly accessible content, such as prohibitions on scraping or on use for AI training, do not make that content any less publicly available. Like reservations under Article 4, they do not apply to training covered by the mechanism (see section 5).

The remuneration is accounted for at one point in the value chain only, by the first company that commercially supplies a model on the EU market. A model provider that sells access to its model, whether through an API, a commercial licence or hosted inference, accounts on its EU revenue from that sale. Companies that build products or services on model access bought in this way have nothing further to account for.

The obligation also arises where a model is used commercially without access to it being sold at arm's length. In that case the value of the model access is deemed. This covers three situations: a company that builds its own model into products or services it supplies to others; a company that runs an open model itself in a product or service it supplies commercially; and services offered at no monetary price (see section 3). Supplies within a group of companies are treated in the same way: what counts is the value at arm's length, not the price charged. Where a company buys model access from a supplier that has not accounted for the remuneration, for example a provider established outside the EU that has not registered, the buyer accounts on what it paid for that access.

The result is that each commercial supply is remunerated at only one point in the value chain, and the total does not depend on how the value chain is organised: an integrated company that builds its own model into the services it supplies pays the same as a chain of separate companies supplying the same service, and a company that runs an open model commercially pays the same as one that buys access to a proprietary model.

Several activities fall outside the trigger because they involve no commercial supply. Research, and the development and release of models under open-source licences, take place before the point at which the obligation arises. The obligation arises when such a model is commercially supplied, for example by a company that sells hosted access to it. Internal use of a model by an organisation, without supplying it or a service based on it to others, is also outside. This is different from a company building its own model into products or services it supplies to others, which is in scope at a deemed value (see above). In addition, companies whose EU revenue from in-scope supplies falls below a threshold are outside the system entirely.

Open design question: how a company that fine-tunes a model it has bought access to, and supplies the result commercially, is treated: as the first supplier of a new model, or as covered by what was accounted for on the base model.

Open design question: where the line runs for open models. Their release falls outside the mechanism, and their commercial supply falls within it. Open is whether the obligation should fall on every company that builds an open model into products it sells, or only on companies that offer access to the model as a service (hosting or API access) and on integrators above a threshold.

2. MECHANISM

The obligation takes the form of a statutory remuneration claim. It cannot be waived, and it is subject to mandatory collective management. Remuneration is paid into a single EU-level fund, which divides it into two shares: one for creators and rightholders, which it passes to collective management organisations for distribution to their members, and one for the stewards of publicly available content, which it allocates itself (see section 4).

Collective management is the only channel that can reach the creators and rightholders this mechanism is meant to remunerate. Most of the content models are trained on cannot be traced to an identifiable rightholder who could negotiate individually, and attributing value to individual works is neither possible nor necessary. Mandatory collective management is also well established in EU copyright law. Cable retransmission rights must already be exercised through collecting societies, and authors and performers already hold an unwaivable right to equitable remuneration for rental. The European Parliament's own-initiative report on copyright and generative AI likewise points to collective management as the channel through which remuneration for AI uses should flow, and the Commission's consultation lists a remuneration right for uses of protected content by generative AI providers among the options it is considering.

A single EU-level fund avoids 27 national schemes with different rules, rates and collection procedures. That would fragment the market for AI providers and rightholders alike, and would weaken the position of beneficiaries in smaller member states.

3. BASIS AND RATE

Remuneration is calculated as a percentage of the EU revenue from the supply of model access, or of its deemed value, as described in section 1. As a general rule a percentage of revenue is proportional: it scales with the value a company captures from the commercial supply of models built on publicly available content, and it falls lightly on companies that are small or still building a market.

The base is the trained model. What the mechanism remunerates is the value a model accrues from training on publicly available content, as realised when it is commercially supplied. What a company adds on top of a model, in its own product, data and customer relationships, is its own contribution. Remunerating that as well would turn the mechanism into a charge on the use of AI in general, rather than a payment for the value taken from publicly available content.

Where the value of model access is deemed, it is set by reference to the price of equivalent model access on the market, expressed as a reference value per unit of use. This removes the need to attribute the revenue of a larger product to the model built into it. A company that buys model access is covered by what its supplier accounted for. A company that uses its own model, or runs an open model itself, accounts on the deemed value of the access it uses, however the product that incorporates it is priced.

Services offered to users at no monetary price, such as free tiers or ad-funded services, are covered by what the supplier accounted for where the model access is bought in. Where it is not, the deemed value of model access is a fixed amount per active user in the EU.

The rate, the per-user amount and the reference values for deemed model access are set by Commission delegated act and reviewed at regular intervals. The market for AI services is developing too quickly for a rate fixed in the legislation itself to remain appropriate for long. Setting it by delegated act keeps it under democratic control while allowing it to follow the market.

Open design questions: the unit of use for deemed model access, and how reference values follow changes in market prices for model access.

4. DISTRIBUTION AND GOVERNANCE

The fund splits the proceeds between two destinations. The larger part goes to creators and rightholders, through collective management organisations. A significant, fixed share goes to the stewards of publicly available content: the institutions that create and preserve the public information resources models are trained on. These include cultural heritage institutions such as libraries, archives and museums, open knowledge projects and public service media.

This is a split at the level of the fund, not a deduction from what rightholders receive. The remuneration is paid for the use of a body of content far wider than the works actively managed by rightholders, and the stewards' share reflects that. The value that training extracts comes in part from content with no identifiable rightholder or whose rightholders do not seek payment, such as public domain material, openly licensed works and works whose rightholders cannot be traced. That part cannot be distributed to individual rightholders. Applying it collectively to sustain the production of such content follows established practice: the Collective Rights Management Directive allows non-distributable amounts to be used for social, cultural and educational purposes, and in some member states part of private copying remuneration is set aside for cultural purposes. The split at the level of the fund means the mechanism does not

need to attribute value to individual works, which cannot be done reliably at the scale at which models are trained. The share is fixed in the instrument, so that it is not renegotiated with each distribution, and is reviewed at regular intervals.

The claim must reach creators, not only the organisations that already hold rights in their work. A guaranteed share of the remuneration for creators and rightholders is therefore reserved for authors and performers, and cannot be assigned or bought out by contract. This share could extend to the entire rightholder share. Creators and rightholders who want to benefit directly join a collective management organisation. Those who do not still benefit indirectly, through the share that goes to sustaining the information ecosystem as a whole.

Allocation across member states follows a key, to be determined.

The fund is governed by a board drawn from three categories of members. The first are collective management organisations, balanced across member states and types of repertoire. The second are representatives of the stewardship sector: cultural heritage institutions, public interest media and organisations that maintain open knowledge resources. The third are independent experts, including academics. Decisions on allocation require a two-thirds majority, with support from each of the three categories. No single group can decide alone, and the fund is neither an extension of the collective management system nor an instrument of government spending.

The mechanism raises remuneration on EU revenue and is designed to sustain the European information ecosystem. It does not claim to settle the wider question of how value from AI should flow to information producers globally.

Open design questions: the size of the guaranteed share for authors and performers; the size of the stewards' share; the key for allocation across member states.

5. RELATIONSHIP WITH ARTICLE 4 AND OTHER COPYRIGHT RULES

The remuneration covers training on publicly available content, and only training. Training means the training of a model by its developer, including any further training of that model by the same developer. It covers the reproductions made for that purpose, such as collecting, cleaning, curating and storing training data. It does not cover reproductions kept for use at the point of inference, such as retrieval indexes. For this training, remuneration takes the place of individual reservations under Article 4 of the CDSM Directive. Reservations do not apply to it, and remuneration falls due for as long as a model trained in this way is commercially supplied on the EU market.

The mechanism complements Article 4 rather than replacing it: everything outside training on publicly available content is unaffected. Uses at the point of inference, such as retrieval-augmented generation and grounding, are also unaffected, and rightholders remain free to

license them individually or collectively. This is where they have real leverage, because these uses depend on continued access to current content. Content that is not publicly available stays fully subject to Article 4 and remains available for licensing, which for rightholders with large catalogues is a significant market in its own right.

Rightholders therefore keep control, but they exercise it through access rather than through reservations. Content placed behind access controls stays out of the mechanism. Content that remains openly accessible earns remuneration through it. This reverses the incentives of the current system. Today a reservation costs nothing and openness earns nothing, so the rational response is to withdraw content from the open web. That withdrawal also cuts off archives, researchers and other public-interest users who rely on the same content. Under the proposed mechanism, openly accessible content is what generates the fund, which gives rightholders collectively a reason to keep the open web open.

The mechanism also has clear limits. It does not make lawful the use of content obtained unlawfully. It does not cover inference-time uses. And it does not affect liability for outputs that infringe copyright.

It can coexist with robust training data transparency obligations. Once remuneration no longer depends on showing which works were used, transparency is no longer needed to enforce payment but it remains important for public understanding of how AI systems are built, and for establishing which models fall within the scope of the mechanism.

Research, and the development and release of models under open-source licences, involve no commercial supply. Their training is covered in the same way, and remuneration falls due only if and when such a model is commercially supplied (see section 1).

6. LEGAL BASIS AND INSTRUMENT

The mechanism should be set out in a new, self-contained instrument, which provides that for training general-purpose AI models on publicly available content, remuneration takes the place of reservations under Article 4 of the CDSM Directive. A targeted amendment to the AI Act would align it with this: the obligation in Article 53(1)(c) to identify and comply with reservations would not extend to training on publicly available content covered by the instrument, and providers of general-purpose AI models that commercially supply such models on the EU market would be required to account for the remuneration.

The legal basis is Article 114 TFEU, the internal market provision on which both the CDSM Directive and the AI Act rest. If member states each answered the question of remuneration for AI training in their own way, the result would be 27 different regimes for AI providers and rightholders alike. Providers would face conflicting obligations, and rightholders in smaller markets would have little bargaining power. A single EU mechanism avoids that fragmentation and gives rightholders collectively a position that no national scheme could.

Because Article 114 does not extend to fiscal provisions, the nature of the scheme matters. The mechanism is a remuneration scheme, not a tax. Its proceeds go to creators, rightholders and the stewards of publicly available content, through an independently governed fund. They do not go into public budgets. EU copyright law already contains comparable schemes. The closest are the rights to equitable remuneration that authors and performers hold for the rental of their works, and that performers and producers hold for the broadcasting and communication to the public of phonograms. These are statutory remuneration rights, commonly managed collectively, and they have long formed part of EU copyright law on an internal market basis. The Court of Justice has held that equitable remuneration is to be assessed in the light of the value of the use in trade (SENA, C-245/00). That is the logic of this mechanism: it remunerates the value that models accrue from training on publicly available content, as realised when they are commercially supplied on the EU market. It does not compensate for harm to individual rightholders, which is how fair compensation for private copying is calculated (Padawan, C-467/08). The mechanism takes no position on whether training on publicly available content requires authorisation. The obligation arises from commercial supply on the EU market, irrespective of where the training took place and whether it was lawful under the law that applied there.

Because the remuneration becomes part of the obligations of providers of general-purpose AI models under the AI Act, it applies to providers established outside the EU on the same terms as to those within it, and is enforced by the AI Office with the powers it already has. Under the allocation set out in section 1, most of the remuneration is accounted for by these providers.

Open design question: whether the same route reaches non-EU companies that do not supply models themselves but offer services built on open models to users in the EU (the AI Act only covers providers of general-purpose AI models).

ABOUT OPEN FUTURE

Open Future is a European think tank that develops new approaches to an open internet that maximize societal benefits of shared data, knowledge and culture.

This paper is published under the terms of the Creative Commons Attribution License.