

# Kluwer Copyright Blog

## Remuneration for use of works in text and data mining

Damian Flisak (Blue Dragon Institute) · Friday, March 28th, 2025

The debate on the use of copyrighted material to train generative AI models is evolving, shifting its focus from whether compensation is due to creators, to determining the structure and specifics of a remuneration system. The discussion seems to be focused on a number of remuneration schemes, three of which deserve special attention. First, an opt-out mechanism is proposed as a way to license (or gain consent for) the use of protected works for AI training (see interesting comments by [J.P. Quintais](#), especially on pp. 13). The second proposal, promoted by [C. Geiger](#) and [V. Iaia](#), is to remove the provision for commercial text and data mining (TDM) in favour of the introduction of a statutory licence for the use of a work to train AI models. A third option, proposed by [M. Senftleben](#), is to make AI-generated content subject to a kind of royalties, paid to the community of creators.

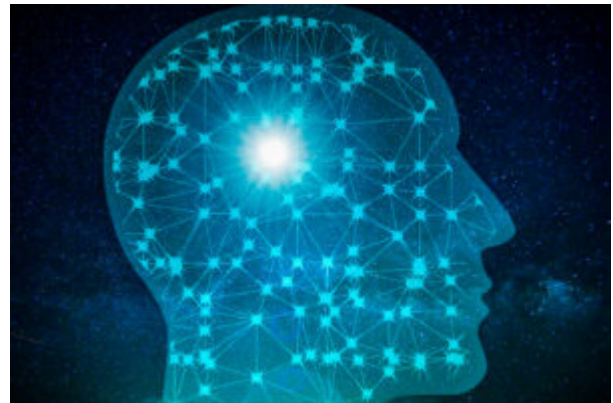


Image by Mohamed Hassan via PxHere

This post will take a closer look at these three proposals.

### Opt-out reservations as a licensing mechanism

Based on existing rules, it is possible to use the opt-out model of Article 4 of the [Copyright in the Digital Single Market \(DSM\) Directive](#) as a starting point for constructing a scheme for the compensation of creators. The logic is simple: the inclusion of an opt-out by the right holder means either a complete lack of consent for use to train AI models, or the granting of consent for use in exchange for a reasonable fee.

### *Opt-out: many challenges*

The primary role of the opt-out is to reserve the right to use protected works as a training substrate for AI models. Despite the passage of time since the entry into force of the DSM Directive in 2019, a technological standard for the provision of the opt-out has still not emerged. In practice, the text file robots.txt, which contains guidelines for robots crawling websites, is used as the primary tool.

At the same time, this file is unable to force crawlers to behave in the way specified in the opt-out. Another problem is who has the right to place a disclaimer and what the legal effects of such a disclaimer are (i.e., whether a disclaimer against a work located in one place on the Internet extends to all other places). Finally, it is problematic that the disclaimer can be placed at website level as in practice it is not always the case that the entity that is entitled to the works has the right or ability to place the relevant disclaimers at the website level.

### *Rights management*

Given the unprecedented scale of data use in the TDM process, it is crucial for efficiency to establish a system for collective management of the “right to train”. One idea worth considering is the creation of Extended Collective Licensing (ECL) for works for use in TDM relating to training of AI models. There is no one single pan-European ECL system, however one is being considered in [Spain](#). Certainly, it should be noted that such a copyright management system represents a certain deviation from the principle of the right holder’s independent ability to use and dispose of his work. It therefore constitutes, to some extent, a restriction of the rights of those entitled to use and dispose of the work. The main purpose of the ECL system is to make the possibility of receiving remuneration for the use of a work more realistic and to strengthen the bargaining position in determining the remuneration due.

The creation of an ECL system in certain situations is permitted by Article 12 of the DSM Directive. Its underlying logic, aligns well with the practicalities of TDM. The amount of data underlying the capacity of these systems is virtually incalculable, the exploitation process itself is extremely complex (and still poorly understood), and the ability to claim individual remuneration is completely illusory (except perhaps for some European or global leaders). It is therefore reasonable to assume that the establishment of an ECL system will be more advantageous to rightholders than attempts to seek remuneration individually.

The systemic argument is also noteworthy: since some countries, such as [Poland](#), have adopted [ECL for the related right of press publishers](#), the rationale for such a system for TDM seems at least as adequate.

### *Applicability*

A key advantage of using the opt-out is that it is based on already existing legislation. Any discussion of a different compensation model (see below) would require significant regulatory change, which is time-consuming. Moreover, its feasibility is uncertain.

The need to base the remuneration system on the opt-out (even taking into account the many practical doubts) has reached the consciousness of EU countries. A summary of the Policy

Questionnaire on the Relationship between Generative Artificial Intelligence and Copyright and Related Rights, prepared during the Hungarian Presidency in the second half of 2024, provides valuable information in this regard. The [document](#) (see pp. 21) indicates that “a pair of Member States” identified the ECL model as optimal, with one country even proposing a mandatory copyright licensing model.

Of course, the introduction of such a system will require the resolution of a number of legal and technological issues. There are also some disadvantages in terms of “transaction costs” incurred by the developers of AI models, related among others to the need to verify the attractiveness of the data package offered and the entitlement to it. However, it is not excluded that also this stage of evaluating the value of an opt-out data package can be automated.

### **Statutory Licence for Commercial TDM**

An [alternative](#) is to replace the commercial TDM with a statutory licence. This would allow the use of the right to train for a fee (“permitted-but-paid”). This proposal is based on an analogy with the private copying exception.

The undoubted advantage of this solution is its simplicity: without the need to introduce exceptions (of the opt-out type), the use of protected works for TDM would be allowed, just as the copying of books is allowed in libraries. However, this solution has fundamental disadvantages. First of all, all the problems mentioned above, either in terms of determining a fair remuneration for the use of work, or in terms of setting up a functioning infrastructure for collective licensing, remain valid under this model. The increasing aversion to further regulations has already been mentioned above.

### **Charging for AI-generated content**

This [proposal](#) takes a different approach. It focuses not on the stage of training generative AI (input), but on the output. According to this idea, companies that use AI systems that generate content that can be a substitute for human works would be charged a payment. The amount of this payment should be calculated based on revenue and it should be distributed to a broad group of creators through a mandatory collective licensing system. One of the cornerstones of this concept would be not so much the payment of money to “AI-affected” creators, but the funding of various initiatives to support the phenomenon of human creation. The theoretical basis for such support for living creators is the concept of *domain public payant* (see [Adolf Dietz](#)’s work). According to this idea, a part of the benefits derived from the exploitation of works whose copyright has expired should accrue to living creators in order to improve their permanent state of underfunding.

The primary benefit of the proposed concept is again its simplicity: it involves the collection of payments for generated content, and eliminates the opaque phase of training AI models. Yet, nearly all the problems identified with establishing a licensing system for the ‘right to train’ still pertain to this. In addition, from a purely copyright point of view, it is problematic to link the

obligation to pay for AI content that does not infringe copyright, but is at the same time competitive with human creativity.

## Conclusion

The above leads to the conclusion that the most pragmatic and realistic solution is to license content using an opt-out mechanism. It is based on existing legislation and the very idea of using this mechanism is beginning to be discussed intensively. A good step in this regard is the creation, under the Polish presidency in the Council of the EU, of a [policy questionnaire](#) on the challenges facing CMOs in the EU Member States. One of the points of this questionnaire is to inform Member States about the practice of licensing rights to works for the purpose of training AI models.

---

*To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please [subscribe here](#).*



2024 Future Ready Lawyer Survey Report

# Legal innovation: Seizing the future or falling behind?

Download your free copy →

 Wolters Kluwer

 Future Ready

**LAWYER**

This entry was posted on Friday, March 28th, 2025 at 11:53 am and is filed under [Artificial Intelligence \(AI\)](#), [European Union](#), [Text and Data Mining \(TDM\)](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.

