ANNALES
SCIENTIA POLITICA

ANDRZEJ MADERA

Academy of Justice in Warsaw
Poland
E-mail: amadera@poczta.onet.pl

Abstract:

The article addresses the lack of adequate legal mechanisms regulating the use of copyrighted works in the training of artificial intelligence (AI) algorithms, particularly in the context of copyright law. The author argues that current licensing solutions, especially those based on collective rights management organizations are insufficient given the scale and nature of AI training operations. In response, the article proposes the adoption of a statutory compensation model inspired by the copy levies system used in the European Union. Under this framework, AI developers would be required to pay a fee for using protected content, with the proceeds distributed by authorized organizations to the rightsholders. The article analyzes both the advantages and limitations of this model, highlighting its potential as a tool for balancing the interests of creators and the innovation sector. The author contends that only a fair and systemic compensation mechanism can ensure the sustainable development of both AI technologies and creative industries.

Keywords:

Artificial intelligence (AI), AI model training, copy levies, copyright law, statutory compensation.

A STATUTORY COMPENSATION MODEL FOR AI TRAINING BASED ON THE COPY LEVIES PAYMENT FRAMEWORK

ANDRZEJ MADERA

Academy of Justice in Warsaw
Poland
E-mail: amadera@poczta.onet.pl

Abstract:

The article addresses the lack of adequate legal mechanisms regulating the use of copyrighted works in the training of artificial intelligence (AI) algorithms, particularly in the context of copyright law. The author argues that current licensing solutions, especially those based on collective rights management organizations are insufficient given the scale and nature of AI training operations. In response, the article proposes the adoption of a statutory compensation model inspired by the copy levies system used in the European Union. Under this framework, AI developers would be required to pay a fee for using protected content, with the proceeds distributed by authorized organizations to the rightsholders. The article analyzes both the advantages and limitations of this model, highlighting its potential as a tool for balancing the interests of creators and the innovation sector. The author contends that only a fair and systemic compensation mechanism can ensure the sustainable development of both AI technologies and creative industries.

Keywords:

Artificial intelligence (AI), AI model training, copy levies, copyright law, statutory compensation.

Introduction

The use of copyrighted works and related subject matter by artificial intelligence (AI) often occurs in violation of copyright law. Such infringements may concern both input data and output content generated by AI algorithms. In particular, with respect to input data, attention must first be paid to the training process of AI systems, which relies on large datasets many of which consist of works protected by copyright. The training of AI involves analyzing and processing vast amounts of data to develop models capable of generating new content or making decisions. These data sources frequently include books, newspaper articles, images, music recordings, photographs, and other forms of creative expression protected under intellectual property law. The absence of appropriate authorization for using such works raises significant legal concerns, especially in the context of potential violations of both the moral and economic rights of authors (Škiljić, 2021, p. 1338–1345).

In both legal scholarship and courtrooms across various jurisdictions, a vigorous debate is ongoing as to the extent to which the use of copyrighted content in AI training constitutes a breach of existing copyright rules (Geiger, 2021, p. 383–394; Margoni & Kretschmer, 2022, p. 685–701; Dermawan, 2023, p. 1–25). The lack of a unified approach makes it necessary to ex-

amine this issue through the lens of different legal systems. In practice, however, the most influential perspectives remain those of the United States and the European Union, as many democratic countries tend to follow the regulatory models developed within these two legal frameworks, particularly in the field of emerging technologies.

In the United States, several high-profile lawsuits are currently underway. Organizations representing creators, as well as individual authors themselves, including the prominent New York Times have initiated legal action against major AI developers such as OpenAI, Midjourney, and Stable Diffusion. These cases concern the alleged unauthorized use of copyrighted materials in the training of AI models without proper licensing or consent from rights holders (Allyn, 2024). Such proceedings illustrate the scale of the issue and highlight the legal uncertainty in this area, which may lead to serious consequences. At the same time, they pressure lawmakers to adapt copyright law to the challenges posed by the rapid development of AI technologies. This issue goes beyond isolated claims and takes on a systemic dimension, as it touches upon the core principles of intellectual property protection in the digital age (Lim, 2023, p. 841–842).

One possible response to this situation is the development of new legal mechanisms most notably, statutory compensation systems inspired by existing levy schemes such as the private copying levies in the European Union. Such mechanisms could offer fair remuneration to creators while enabling the continued growth of AI-based technologies.

The following sections of the article present a proposal for the introduction of a statutory compensation model for AI training based on copyrighted works and related rights, inspired by the legal structure of private copying levies. The research questions addressed in the article concern the feasibility of adapting current legal frameworks surrounding copy levy payments to solve the problem of compensating creators for the financial losses incurred as a result of the use of their works and related subject matter in AI training.

The aim of the article is to identify the potential benefits and risks associated with such a model and to propose the legal modifications necessary to adapt the current reprographic levy system to contemporary technological conditions. The article adopts a doctrinal-analytical approach, focusing on the examination, interpretation, and practical application of existing legal provisions. It employs normative analysis of copyright law regulations and doctrinal methods involving the examination of scholarly positions and legal literature.

The AI Training Process in the Light of EU Law

Currently, the law of the European Union does not comprehensively or directly regulate the use of copyrighted works in the process of training artificial intelligence (AI) algorithms. The Artificial Intelligence Act (AI Act), while a landmark legislative instrument setting the framework for the functioning of AI in Europe, does not explicitly resolve the issue of copyright infringement in this context. The provisions of the AI Act regarding the training of AI models are limited to two obligations imposed on providers of general-purpose AI models:

- the obligation to comply with existing copyright law, in particular Article 4 of Directive 2019/790 (see Article 52c(1)(c) of the AI Act); and
- the obligation to prepare and publicly disclose a detailed summary of the content used in the training of AI models, in accordance with a template provided by the AI Office.

It is important to note that these provisions are found in Chapter II of the AI Act, which concerns the "Obligations of providers of general-purpose AI models", and not in a separate section dedicated to copyright law, since no such section exists in the Act. Moreover, as stated in the final sentence of Recital 60ka of the AI Act: "This Regulation shall not affect the enforcement of copyright rules under Union law." This means that the AI Act does not create new copyright norms and cannot serve as a basis for interpreting existing copyright provisions. In the EU legal framework, the rules applicable to AI training are instead found in Articles 3 and 4 of Directive (EU) 2019/790. This directive introduces rules on so-called text and data mining (TDM), which refers to the automated analysis of large datasets in order to extract information (Lemley, 2024, p. 190–200).

Directive 2019/790 provides for: a research exception in Article 3, which permits TDM by research organizations and cultural heritage institutions without the need to obtain authorization from rights holders; and a commercial exception in Article 4, which allows TDM for commercial purposes, subject to the condition that rights holders have not opted out via an appropriate declaration. In practice, this means that if the rights holder has not issued an opt-out declaration, AI developers may legally and freely use publicly available works for training purposes. However, if an opt-out has been exercised, a license must be obtained and appropriate remuneration paid (Rosati, 2019, p. 198–210; Margoni & Kretschmer, 2022, p. 685–695). The definition of TDM is set out in Article 2 of Directive 2019/790 and refers to a set of methods and techniques for extracting useful information from document collections by identifying and using patterns from unstructured textual datasets. From a formal standpoint, TDM is limited to generating new information based on a corpus of documents, which can then undergo processes such as text categorization, clustering, concept extraction, sentiment analysis, or modeling relationships between elements (Truyens & Van Eecke, 2014, p. 153–170). The underlying purpose of TDM is to reduce the effort required to search through massive datasets of any type that meet the criteria of Big Data (Shemilt et al., 2014, p. 31–49). It is important to emphasize that the notion of an AI system is broader and goes far beyond the purpose of TDM, which is only to generate information containing patterns, trends, and correlations. What distinguishes AI systems from TDM is their ability to generate entirely new content, texts, images, or other media, which did not previously exist. In this sense, AI systems can impart a unique expressive form to the extracted information, resulting in the creation of new, original outputs (Floridi & Chiaraatti, 2020, p. 681–694). The key distinction lies in the fact that TDM alone does not generate new creative works, but merely extracts information, which may serve as a starting point for the subsequent production of texts, films, or other media (Tylec et al., 2024, p. 5–13; Zewe, 2023).

Although the DSM Directive provides a legal framework for the lawful use of protected materials in AI training, it does not fully resolve the issue of appropriate remuneration for creators. When rights holders exercise the opt-out clause, a license must be obtained on market terms, but the practical implementation of this model poses significant challenges. Traditional licensing systems operated by collective management organizations are ill-suited to the specific nature of AI training, which involves analyzing hundreds of millions of works—often in an automated and large-scale manner (Geiger et al., 2018a, p. 95–112; Geiger et al., 2018b, p. 814–844).

The Traditional Licensing System as an Inadequate Solution for AI Training

The traditional licensing system for works, operated through collective management organizations, was designed to regulate relatively limited and individual access to protected content. It is based on the premise that the use of a work is concrete, identifiable, and attributable to a specific party and context. Each instance of use, whether it be playing music, publicly performing a literary work, or reproducing an image, can be clearly identified, documented, and covered by a relevant license entitling the rightsholder to specific remuneration. The nature of training artificial intelligence (AI) models fundamentally differs from this traditional scenario. AI training involves processing and analyzing vast datasets, which often include hundreds of millions of copyright-protected works (Korngiebel & Mooney, 2021, p. 1). This process is massive, automated, and executed on a bulk scale, which gives rise to several significant legal and organizational challenges (Senftleben et al., 2022, p. 67–86). The primary issue is the inability to individually identify the works and related subject matter used in AI training. In the case of traditional licenses, it is possible to determine which specific work is being used and in what context. During AI training, however, input data is analyzed in aggregate, and individual works often lose their distinctiveness during processing, making it impossible to monitor and license usage at an individual level. The second problem is that traditional collective management organizations are not equipped to handle transactions involving hundreds of millions of works simultaneously. From both technical and administrative standpoints, negotiating

separate licenses for each potentially used work is unfeasible. A third essential aspect of the phenomenon is that AI training is conducted without human involvement at the content analysis stage. This process relies on automatic "reading" of data by algorithms, without the conscious act of use characteristic of traditional uses of works. As a result, the classical model, built on conscious, traceable exploitation, becomes obsolete. Furthermore, traditional licensing models depend on usage reporting and remuneration schemes that are not suited to the aggregate and statistical nature of AI training. Existing mechanisms do not provide tools to effectively calculate remuneration based on the actual scope of work usage in training. Consequently, the traditional licensing system proves inadequate in the context of AI technologies. In theory, it would be possible to require AI model developers to obtain licenses before training begins. However, as outlined above, such a system would be practically unworkable (Samuelson, 2023, p. 159–161; Senftleben, 2024, p. 1–28).

The rise of generative AI models has thus necessitated a reassessment of the adequacy of traditional copyright licensing regimes. The conventional approach, which relies on individual licenses granted through collective management organizations, is not adapted to the scale and automation characteristic of AI training, which involves bulk analysis of hundreds of millions of works, often in fragmentary form (Geiger & Iaia, 2024, p. 1–24). In response, various legislative and technical models have been proposed to reconcile the interests of creators with the advancement of technology (Sag, 2023, p. 316–321).

One prominent proposal is the introduction of a so-called statutory license. This solution, supported by scholars such as Christophe Geiger and Valentino Iaia (Geiger & Iaia, 2024, p. 1–24), would allow AI developers to use copyrighted works without individual prior consent. After use, the works would be identified, remuneration would be calculated, and payments would be distributed to rightsholders, with the process potentially managed by collective management organizations or specialized copyright institutions. Geiger and Iaia emphasize that this construction would respect creators' rights while eliminating the unrealis-

tic burden of obtaining licenses for each individual work.

An alternative model is the implementation of an Extended Collective Licensing (ECL) scheme, already functional in certain legal systems such as in the Nordic countries. Spain, for instance, has proposed adopting an ECL model for training general-purpose AI models (Nobre, 2024). Under this framework, collective management organizations would be authorized to grant licenses on behalf of all authors, including non-members, while authors would retain the right to opt out. ECL simplifies the licensing process and reduces transaction costs, making it an attractive tool for regulating large-scale data usage in AI training. A similar solution was introduced in the United Kingdom in 2025, allowing for legal and transparent use of protected content while ensuring financial compensation for authors without requiring individual negotiations (The Guardian, 2025).

In the United States, the "Generative AI Copyright Disclosure Act" is currently under consideration. This bill would require AI developers to report to the Copyright Office the list of works used in model training at least 30 days before public release. While the proposed legislation does not create a remuneration system, its purpose is to increase transparency and empower authors to monitor the use of their works (Kline, 2024).

Another proposed solution involves developing a system in which creators receive compensation proportional to the contribution of their works to AI training. This would require implementing a mechanism for tracking the influence of specific works on AI model performance and automating the compensation allocation accordingly. While technologically ambitious, such a system would necessitate sophisticated methods for analyzing the contribution of individual data to machine learning processes (Wang et al., 2024; Senftleben, 2024, p. 1–28).

At present, none of the above proposals offers a perfect solution. However, statutory licenses, extended collective licensing, and disclosure-based mechanisms appear better suited to address the challenges of the large-scale, automated nature of AI training than the traditional model. Future research should focus on developing effective and equitable compensation systems that reconcile rapid technological

advancement with robust copyright protection. In the author's view, the concept of copy levies payment holds significant potential in this context. Under such a model, AI developers would pay a fee as compensation for the use of copyrighted works in model training, analogous to the payments made by manufacturers of copying devices under private copying schemes. This analogy suggests that adapting the reprographic levy system could be a viable solution for licensing protected works in the context of AI training. However, this approach would require careful tailoring to the specificities of AI technologies and the establishment of appropriate control and distribution mechanisms.

Proposal for a New Solution: A Statutory Compensation System

Given the challenges outlined above in adapting the existing licensing framework to new technological conditions, the implementation of a statutory compensation system modeled on the legal framework of private copying levies deserves serious consideration. This system, which operates in many European countries, imposes levies on manufacturers, importers, and distributors of devices and media that enable the copying of works, in order to compensate creators for losses stemming from private copying of their works. Similarly, in the context of AI, fees could be levied on AI model developers or service providers that use protected works to train algorithms. These funds could then be distributed to rights holders by specialized organizations, ensuring fair remuneration and balancing the interests of users and rights owners. The private copying levy system constitutes one of the pillars of the European copyright protection regime. It entails an additional fee imposed on manufacturers, importers, and distributors of specific devices and storage media capable of copying legally protected works. The reprographic levy can be described as a compensatory payment intended to remunerate copyright and related rights holders for lost potential income resulting from the private copying of their works. The mechanism applies to devices such as printers, scanners, smartphones, tablets, computers, DVD burners, and storage media like CDs/DVDs, USB drives, memory cards, hard drives, and servers (Oksanen & Välimäki, 2005, p. 25–30;

Netanel, 2003, pp. 5–25; Fisher, 2004, pp. 199–210). Generally, the fee is collected from entities placing these products on the market (manufacturers, importers, vendors), and subsequently redistributed through collective rights management organizations to the entitled parties. The reprographic levy system supplements the legal framework of the so-called private use exception. This exception allows individuals to make copies of works and related subject matter for personal use, meaning they are not required to obtain the author's permission each time. In exchange, rights holders receive financial compensation. One of the central goals of this solution is to strike a balance between the rights of authors and public interests. Users can legally copy works for personal purposes, while rights holders are guaranteed fair compensation. The legal basis for the EU's copy levy system is found in Directive 2001/29/EC. Article 2 of this directive guarantees authors the exclusive right to authorize or prohibit reproduction of their works, while Article 5(2)(b) allows Member States to introduce exceptions for private copying provided that rights holders receive fair compensation. While this system enables access to content, supporting education, culture, and innovation, it also safeguards the financial interests of authors and their successors. However, it is important to recognize that the mechanism is designed to compensate for losses resulting from legal copying; it does not address unauthorized uses. It was created to balance the financial losses suffered by authors and rights holders due to lawful private copying of their works in non-commercial settings (Hugenholtz, 2012, p. 184–200). Currently, the system is operational in 25 out of 27 EU Member States, though its forms vary significantly in terms of rates, scope, collection, and distribution procedures.

As previously stated, the author believes that adapting the concept of the copy levy system to the training of AI models is a promising avenue. This solution, however, would need modification to function effectively under the technological realities of AI. One notable advantage is administrative simplicity: the reprographic levy system is relatively easy to implement and manage, which may facilitate its adaptation to AI needs. Collective rights management enables efficient handling of rights through organizations, which benefits creators

lacking resources for individual negotiations. However, a drawback of the reprographic levy model is its inability to precisely determine which works were used in AI training, potentially leading to unequal remuneration. Yet this limitation aligns with the very nature of AI training processes, which inherently lack transparency regarding the specific works used. The absence of precise tracking may discourage creators from making their works available, fearing a loss of control. Additionally, without proper oversight mechanisms, such a system could be vulnerable to abuse and inefficient fund allocation. This solution could be enhanced by incorporating a proportional remuneration model, where creators are compensated based on the contribution of their works to AI training. As noted above, this would require technological systems capable of tracking and evaluating the influence of individual works on AI model performance, allowing for more accurate attribution of payments to specific authors.

Traditional reprographic levies apply to physical media and copying devices. In the context of AI, these should be extended to cover training processes involving copyrighted content. This would mean imposing levies on companies developing AI models, proportionate to the extent of protected material used in training. To effectively adapt the copy levy framework to the AI context, it is essential to account for the unique characteristics of AI model training and establish suitable oversight and fund distribution mechanisms. The first necessary step appears to be implementing a requirement for transparency in training data. AI developers should be mandated to disclose information about the sources of data used for training in a way that enables identification of the rights holders. This would enhance transparency and allow creators to monitor the use of their works. A good example of such an initiative is the aforementioned "Generative AI Copyright Disclosure Act" proposed in the United States, which would require AI companies to report to the Copyright Office any copyrighted works used for AI training (Jernite, 2023; Warso et al., 2024, p. 1–3; Senftleben, 2024, p. 1–28).

Another recommended step is the creation of a central registry of works used in AI training. Such a registry would allow creators to register their works and monitor their use, improving copyright management and the distribution of levy funds (Ziaja, 2024, p. 453–459; Keller & Warso, 2023; Senftleben, 2024, p. 1–28). Furthermore, it would be beneficial to establish a specialized organization or authority responsible for managing the distribution of funds collected from AI developers. This body could function similarly to existing collective rights management organizations, ensuring equitable allocation of resources. In the author's view, implementing the measures outlined above would allow for the effective adaptation of the copy levy system to the context of AI, ensuring protection for creators' rights while supporting the development of innovative technologies.

Conclusion

The ongoing debate regarding the relationship between copyright and the development of artificial intelligence increasingly centers on a critical question: Does the obligation to remunerate creators for the use of their works by AI systems hinder innovation? In the author's view, this argument deserves critical reconsideration. On the contrary, a balanced, fair, and rationally established compensation system for the use of creative works may contribute to genuine innovation growth in both the AI development sector and the creative industries. Well-designed copyright frameworks are not a barrier to innovation but a prerequisite for its sustainable development. Protecting authors' rights and ensuring their participation in revenues generated from their works creates incentives for continued creative output — without which AI systems, which rely on source content, would lack the necessary material to enhance their models. The absence of such regulation threatens to erode the "creative ecosystem," which is the foundation of culturally driven technology (Geiger & Iaia, 2024, p. 1–9). Moreover, implementing a fair remuneration mechanism reduces legal risk for technology companies. AI developers operating in a gray zone of legal uncertainty face high litigation costs and the risk of damages for copyright infringement. Establishing transparent licensing and compensation rules would provide a more stable business environment, promoting long-term investment and innova-

tion. In addition, copyright law serves as an instrument of cultural policy, aimed at balancing the interests of creators and users. Relying solely on free, unregulated access to creative content for AI systems could lead to a crisis in the creative sector and, over time, to a depletion of the resources on which machine learning technologies depend. Such a scenario not only threatens cultural diversity but also undermines the sustainable growth of the innovation ecosystem. Balanced copyright protection in the context of AI is not only a market necessity but also a response to the core values of the European Union, such as the protection of creativity, the dignity of labor, and fair compensation. Rational compensation mechanisms for creators whose works are used in AI training are not an obstacle to innovation, rather they are a condition for its sustained and equitable development in both the technology and creative sectors. The need for change is further supported by the many legislative proposals currently under discussion. In the EU, these include: mandatory disclosure of training content used for general-purpose AI models, strengthening the opt-out mechanism under Article 4 of the DSM Directive, potentially through automated tools for rights exclusion, proposals to introduce new regulations for "legal TDM" in commercial data use by AI, with a focus on fair compensation for creators even without active opt-out declarations. Consultations are underway in the European Parliament and Commission on a new sector-specific act concerning data used by AI, which could establish a mandatory compensation system akin to reprographic levies (statutory remuneration schemes). As this overview shows, all current legal developments aim to: increase transparency in the use of protected content, ensure fair compensation for authors, minimize legal risks for AI developers, and maintain a balance between innovation and intellectual property protection. The dynamic pace of legislative reform in this field indicates that lawmakers across jurisdictions have recognized the urgent need to adapt legal frameworks to the technological realities of AI. This paper aligns with those efforts by advocating for a new legal approach based on the well-established copy levies payment model.

References

AI Act – Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), PE/24/2024/REV/1. *Official Journal of the European Union, L 2024/1689*, 12 July 2024.

Allyn, B. (2024, August 16). *New York Times considers legal action against OpenAI as copyright tensions swirl*. NPR. https://www.npr.org/2023/08/16/119420 2562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl

Dermawan, A. (2023). Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese 'nonenjoyment' purposes? *The Journal of World Intellectual Property, 26*, 1–25. https://doi.org/10.1111/jwip.12285

Directive 2001/29/EC. (2001). Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc Directive). *Official Journal of the European Communities, L 167*, 10–19.

Directive 2019/790. (2019). Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. *Official Journal of the European Union, L 130*, 92–125.

Fisher, W. W. (2004). An alternative compensation system. In W. W. Fisher (Ed.), *Promises to keep: Technology, law, and the future of entertainment* (pp. [vložiť strany]). Stanford University Press.

Floridi, L., & Chiariatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds & Machines, 30*, 681–694. https://doi.org/10.1007/s11023-020-09548-1

Geiger, C. (2021). The missing goal-scorers in the artificial intelligence team: Of big data, the fundamental right to research and the failed text and data mining limitations in the CDSM Directive. In M. Senftleben, J. Poort, M. van Eechoud, S. Gompel & N. Helberger (Eds.), *Intellectual property and sports: Essays in honour of P. Bernt Hugenholtz* (pp. 383–394). Kluwer International.

Geiger, C., Frosio, G., & Bulayenko, O. (2018a). Crafting a text and data mining exception for machine learning and big data in the digital single market. In X. Seuba, C. Geiger, & J. Penin (Eds.), *Intellectual property and digital trade in the age of artificial intelligence and big data: Global perspectives and challenges for the intellectual property system* (Issue No. 5, pp. 95–112). Geneva/Strasbourg.

Geiger, C., Frosio, G., & Bulayenko, O. (2018b). Text and data mining in the proposed copyright reform: Making the EU ready for an age of big data? *International Review of Intellectual Property and Competition Law, 49*, 814–844. https://doi.org/10.1007/s40319-018-0722-2

Geiger, C., & Iaia, V. (2024). The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI. *Computer Law & Security Review, 52*, 1–9. https://doi.org/10.1016/j.clsr.2023.105925

Hugenholtz, P. B. (2012). The story of the tape recorder and the history of copyright levies. In B. Sherman & L. Wiseman (Eds.), *Copyright and the challenge of the new*. Wolters Kluwer.

Jernite, Y. (2023). Training data transparency in AI: Tools, trends, and policy recommendations. Hugging Face. https://huggingface.co/blog/yjernite/data-transparency

Keller, P., & Warso, Z. (2023). Defining best practices of opting out of ML training. Open Future. https://openfuture.eu/publication/defining-best-practices-for-opting-out-of-ml-training/

Kline, D. (2024). The Generative AI Copyright Disclosure Act of 2024: Balancing innovation and IP rights. *The National Law Review, 15*(120). https://natlawreview.com/article/generative-ai-copyright-disclosure-act-2024-balancing-innovation-and-ip-rights

Korngiebel, D. M., & Mooney, S. D. (2021). Considering the possibilities and pitfalls of generative pre-trained transformer 3 (GPT-3) in healthcare delivery. *NPJ Digital Medicine, 4*, 1–3.

Lemley, M. (2024). How generative AI turns copyright upside down. *Columbia Science & Technology Law Review, 25*(2), 190–212.

Lim, D. (2023). Generative AI and copyright: Principles, priorities and practicalities. *Journal of Intellectual Property Law & Practice, 18*, 841–842.

Lucchi, N. (2023). ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation, 15*(3), 602–624. https://doi.org/10.1017/err.2023.59

Margoni, T., & Kretschmer, M. (2022). A deeper look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology. *GRUR International, 71*, 685–701.

Netanel, N. (2003). Impose a noncommercial use levy to allow free peer-to-peer file sharing. *Harvard Journal of Law & Technology, 17*(1), 2–83. https://jolt.law.harvard.edu/articles/pdf/v17/17HarvJLTech001.pdf

Nobre, T. (2024, December 11). A first look at the Spanish proposal to introduce ECL for AI training. *Kluwer Copyright Blog.* https://copyrightblog.kluweriplaw.com/2024/12/11/a-first-look-at-the-spanish-proposal-to-introduce-ecl-for-ai-training/

Oksanen, V., & Välimäki, M. (2005). Copyright levies as an alternative compensation method for recording artists and technological development. *Review of Economic Research on Copyright Issues, 2*(2), 25–39.

Rosati, E. (2019). Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. *Asia Pacific Law Review, 27*, 198–217.

Sag, M. (2023). Copyright safety for generative AI. *Houston Law Review, 61*, 295, 316–321.

Samuelson, P. (2023). Generative AI meets copyright. *Science, 381*(6654), 158–161. https://doi.org/10.1126/science.adi0656

Senftleben, M. (2024). AI Act and author remuneration – A model for other regions? *SSRN*, 1–28. https://doi.org/10.2139/ssrn.4740268

Senftleben, M., Margoni, T., et al. (2022). Ensuring the visibility and accessibility of European creative content on the world market: The need for copyright data improvement in the light of new technologies and the opportunity arising from Article 17 of the CDSM Directive. *JIPITEC, 13*, 67–86.

Shemilt, I., Simon, A., et al. (2014). Pinpointing needles in giant haystacks: Use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods, 5*, 31–49. https://doi.org/10.1002/jrsm.1093

Strowel, A. (2023). ChatGPT and generative AI tools: Theft of intellectual labor? *IIC, 54*, 491–494. https://doi.org/10.1007/s40319-023-01321-y

Škiljić, A. (2021). When art meets technology or vice versa: Key challenges at the crossroads of AI-generated artworks and copyright law. *IIC International Review of Intellectual Property and Competition Law, 10*, 1338–6139. https://doi.org/10.1007/s40319-021-01119-w

The Guardian. (2025, April 23). Collective licence to ensure UK authors get paid for works used to train AI. https://www.theguardian.com/books/2025/apr/23/collective-licence-to-ensure-uk-authors-get-paid-for-works-used-to-train-ai

Truyens, M., & Van Eecke, P. (2014). Legal aspects of text mining. *Computer Law & Security Review, 30*, 153–170.

https://doi.org/10.1016/j.clsr.2014.01.009

Tylec, G., Kwiecień, S., et al. (2024). Is it possible to license works used in the learning process of artificial intelligence algorithms? *SSRN*, 2–19. https://doi.org/10.2139/ssrn.4729495

Wang, J. T., Deng, Z., et al. (2024). An economic solution to copyright challenges of generative AI. *arXiv*. https://doi.org/10.48550/arXiv.2404.13964

Warso, Z., Gahntz, M., & Keller, P. (2024). Sufficiently detailed? A proposal for implementing the AI Act's training data transparency requirements for GPAI. Open Future. https://openfuture.eu/wp-content/uploads/2024/06/240618AItransparency_template_requirements-2.pdf

Zewe, A. (n.d.). Explained: Generative AI. MIT Schwarzman College of Computing. https://computing.mit.edu/news/explained-generative-ai/

Ziaja, G. M. (2024). The text and data mining opt-out in Article 4(3) CDSMD: Adequate veto right for rightholders or a suffocating blanket for European artificial intelligence innovations? *Journal of Intellectual Property Law & Practice, 10*, 453–459.