LITTERA SCRIPTA

Economics

Management

Corporate Finance

Finance and Valuation

2/2025

VŠTE

Littera Scripta

(Economics, Management, Corporate Finance, Finance and Valuation)

Ing. Jakub HORÁK, MBA, PhD. (Editor-in-chief)

Address Editor:

Institute of Technology and Business in České Budějovice

Okružní 517/10

370 01 České Budějovice, Czech Republic

Tel.: +420 387 842 183

e-mail: journal@littera-scripta.com

ISSN 1805-9112 (Online)

Date of issue: December 2025

Periodicity: Twice a year Since 2010

The Journal is indexed in:

- ERIH PLUS (European Reference Index for the Humanities and Social Sciences) – in 2015
- CEJSH (Central European Journal of Social Sciences and Humanities) – in 2015
- EZB (Elektronische Zeitschriftenbibliothek) – in 2017
- GOOGLE SCHOLAR – in 2017
- DOAJ (Directory of Open Access Journals) – in 2019

Content

A statutory model of compensation for the use of works in artificial intelligence training as a proposal for systemic strengthening of copyright in the era of machine learning

Andrzej Madera[1]

[1]Academy of Justice in Warsaw, Faculty of Security and Legal Sciences

Abstract

The dynamic development of generative artificial intelligence has revealed the structural inadequacy of traditional copyright mechanisms for the mass, automated processing of data used in model training processes. Classic licensing systems, based on identifiable and individual access to a work, are unable to function in an environment where a single algorithm processes millions of protected content in a way that cannot be reproduced using existing regulatory tools. The article analyzes the legal and economic rationale behind the proposal to introduce a statutory compensation model for the use of works in artificial intelligence training, inspired by the European system of reprographic fees. The study attempts to demonstrate that this model can be an effective tool for balancing the interests of creators and technology companies, while strengthening the stability of the copyright system. It also points to the need to supplement future regulations with data transparency obligations and appropriate institutional instruments to enable the efficient redistribution of remuneration. The results of the analysis lead to the conclusion that statutory compensation can become the foundation for the sustainable development of the creative and technological ecosystem in an era of growing AI autonomy.

Keywords: Artificial intelligence; copyright, machine learning, reprographic fees, statutory compensation, text and data mining, generative AI models, protection of creators.

Introduction

Over the last decade, the process of training artificial intelligence systems has undergone a rapid transformation, primarily due to the development of generative models capable of processing, analyzing, and recombining huge data sets. This data covers a wide range of works, such as literary texts, press articles, photographs, images, and music recordings, most of which are protected by copyright. The current regulatory framework does not provide for mechanisms to adequately regulate the use of such extensive and diverse material, creating uncertainty for both creators and AI developers. The material contained in the submitted document clearly emphasizes that the fundamental problems concern not only the scale of the data, but also the way it is processed, which does not allow for the identification of individual works used in the algorithm training process.

The inability to link a specific work to its use in the training process makes it impossible to apply traditional licensing models designed for the precise and traceable use of materials. This issue is widely discussed in the literature on automatic data mining. Lemley (2024) notes that text and data mining (TDM) processes in the context of machine learning are radically different from classic copying and thus challenge existing concepts of exhaustion of rights. Geiger, Frosio, and Bulayenko (2018a; 2018b) emphasize that the TDM exemptions provided for in European Union law have been designed too narrowly to cover the mass use of content by AI systems. For this reason, it is necessary to create new legal solutions that are adequate to the technological realities.

Training AI models as a challenge to existing copyright mechanisms

The process of training AI models is based on the analysis of huge and heterogeneous data sets which, as the source document aptly points out, include millions of works processed in a way that cannot be traced using the methods used in classic licensing

Floridi and Chiariatti (2020) point out that a characteristic feature of generative models is their ability to create new, previously non-existent content thanks to their ability to extract and transform patterns from input data. This differs radically from the traditional reception of a work, in which exploitation is recognizable and its scope can be verified.

In European copyright law, the 2019 DSM Directive plays a special role, introducing two important regulations concerning text and data mining. Article 3 allows TDM for research purposes, while Article 4 provides for the possibility of TDM for commercial purposes, unless the rightholder objects. These regulations, although groundbreaking at the time of their adoption, do not solve the problem of mass training of AI systems, as they assume the possibility of an effective opt-out by the creator, which in practice is often unfeasible in an environment with billions of data points (Margoni & Kretschmer 2022). Truyens and Van Eecke (2014) point out that TDM as a data analysis method was not designed for large-scale generative machine learning, but as an information extraction technique. As a result, current legal regulations are not compatible with the way modern AI systems use data.

Similarly, practical problems related to collective licensing are raised in the literature on the subject. Senftleben et al. (2022) point out that collective management organizations do not have the tools to control or license the use of hundreds of millions of works simultaneously, which makes traditional licensing models structurally inefficient. The document clearly shows that any attempt to adapt the classic licensing system to the needs of AI training would lead to paralysis on both the technological and legal sides, and that any attempt at individual data licensing is practically impossible.

Inspiration from the European system of reprographic fees and the need for statutory compensation

The reprographic fee system, widely discussed in the literature on private copying (Oksanen & Välimäki 2005; Netanel 2003; Hugenholtz 2012), is one of the most successful examples of a legal tool that reconciles the mass use of works with the interests of creators. It allows for the legal copying of works for personal use, while providing creators with indirect remuneration collected from manufacturers of copying devices. Fisher (2004) notes that the strength of compensation systems lies in moving away from attempts to track specific acts of copying and focusing on compensating for the economic effects of using someone else's work.

The conclusion in the submitted document that a similar model can be applied to AI training is strongly justified by the system

In both cases, we are dealing with mass exploitation of content, which is technically impossible to monitor and license individually. Adapting the idea of private copying levies to AI training processes could therefore involve establishing a statutory levy on entities developing AI models, with the revenue redistributed to creators through a collective management organization or a specialized public institution. Geiger and Iaia (2024) argue that statutory compensation is the most realistic way to ensure remuneration for creators when the identification of the works used is technically unfeasible.

The importance of data transparency and the need for institutional support

The literature also points to the need to introduce solutions that increase the transparency of AI training processes. Jernite (2023) emphasizes that the obligation to disclose training data sources is a minimum requirement for the functioning of any efficient compensation model. The source document also notes that fees alone are not sufficient without the creation of a register of data used in model training and an institution capable of managing the flow of funds.

These solutions would contribute to increasing the trust of creators, reducing information asymmetry, and improving the quality of remuneration redistribution.

Conclusion

An analysis of copyright regulations and a review of the literature clearly indicate that the current system of licensing works is not adapted to artificial intelligence technology, whose fundamental feature is the mass, unidentifiable, and automatic use of data. The statutory compensation model, inspired by the reprographic fee system, is a promising proposal for resolving the conflict between the interests of creators and the need for further development of AI technology. This model is consistent with the European tradition of copyright protection, while at the same time being implementable without the need for radical transformation of existing institutional structures.

The literature clearly confirms that the lack of adequate legal solutions threatens both the destabilization of the creative ecosystem and an increase in legal risks for the technology sector. Therefore, statutory compensation appears not as an obstacle to development, but as

a condition for the sustainable coexistence of creativity and innovation in the future. The source document also emphasizes that the introduction of such a mechanism could reduce the level of legal disputes and strengthen the foundations of a market based on respect for creators' rights and transparency of technological processes.

References

Floridi, L., & Chiariatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds & Machines, 30*, 681–694.

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. *Intellectual Property and Sports: Essays in Honor of P. Bernt Hugenholtz, Kluwer Law International*, 383–394.

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.

Geiger, C., Frosio, G., & Bulayenko, O. (2018). Crafting a Text and Data Mining Exception for Machine Learning and Big Data. *Intellectual Property and Digital Trade, Geneva/Strasbourg*, 95–112.

Geiger, C., Frosio, G., & Bulayenko, O. (2018). Text and Data Mining in the Proposed Copyright Reform. *International Review of Intellectual Property and Competition Law, 49*, 814–844.

Hugenholtz, P. B. (2012). The Story of the Tape Recorder and the History of Copyright Levies. *Copyright and the Challenge of the New, Wolters Kluwer*.

Lemley, M. (2024). How Generative AI Turns Copyright Upside Down. *Columbia Science & Technology Law Review, 25*(2), 190–212.

Lucchi, N. (2023). ChatGPT: A Case Study on Copyright Challenges for Generative AI Systems. *European Journal of Risk Regulation, 15*(3), 602–624.

Margoni, T., & Kretschmer, M. (2022). A Deeper Look Into EU TDM Exceptions. *GRUR International, 71*, 685–701.

Netanel, N. (2003). Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer Sharing. *Harvard Journal of Law & Technology, 17*(1), 2–83.

Oksanen, V., & Välimäki, M. (2005). Copyright Levies as an Alternative Compensation Method. *Review of Economic Research on Copyright Issues, 2*(2), 25–39.

Senftleben, M., Margoni, T., et al. (2022). Ensuring Visibility and Accessibility of European Creative Content. *JIPITEC, 13*, 67–86.

Truyens, M., & Van Eecke, P. (2014). Legal Aspects of Text Mining. *Computer Law & Security Review, 30*, 153–170.

Wang, J. T., Deng, Z., et al. (2024). An Economic Solution to Copyright Challenges of Generative AI. *arXiv*.

Contact address of the author:

Andrzej Madera, Academy of Justice in Warsaw, Faculty of Security and Legal Sciences, Wisniowa 50, 02-520 Warsav, Poland, e-mail: amadera@poczta.onet.pl