

Legal limitations and justifications in the training and deployment of generative AI models under copyright law

Rastislav Funta

Danubius University
Richterova 1171
Sládkovičovo 925 21, Slovak Republic
rastislav.funta@vsdanubius.sk

Denisa Krásna

Bratislava University of Economics and Management
Furdekova 16
Bratislava 851 04, Slovak Republic
denisa.krasna.k@gmail.com

Abstract: The article investigates the legal limitations and exceptions relevant to the training and deployment of generative AI models under EU copyright law. It analyzes whether existing copyright exceptions, such as those for temporary reproduction, scientific text and data mining (TDM) and commercial TDM, can be applied to AI training processes. These exceptions generally do not apply due to fundamental technological and conceptual differences between traditional TDM and AI model training. The authors also evaluate the applicability of other exceptions (e.g., quotation, parody or private use) to the public use and output of generative AI. Ultimately, the article argues that current legal frameworks inadequately address the unique challenges posed by generative AI technologies.

Keywords: *copyright law, generative AI, text and data mining (TDM)*

1. Introduction

The rapid development of generative artificial intelligence (GenAI) models has raised complex legal questions, particularly in the context of European Union copyright law. As these models rely on vast datasets, often containing copyrighted material, for training and generation, the applicability of existing legal exceptions and limitations becomes a critical issue. While provisions such as those for temporary reproduction, scientific text and data mining (TDM) and commercial TDM exist, their suitability for AI training remains uncertain due to significant technological and conceptual differences. Furthermore, the legal treatment of AI-generated outputs and the public accessibility of such models introduce additional challenges. In light of these concerns, this article investigates the following research question: To what extent do current EU copyright limitations and exceptions adequately address the legal challenges posed by the training and deployment of generative AI models?

2. Objective and methodology

The objective of this article is to evaluate whether current EU copyright limitations, particularly those related to TDM, are adequate for the legal treatment of generative AI model training and deployment. The research applies a combination of doctrinal legal analysis, systematic interpretation and teleological interpretation of EU copyright law. Additionally, comparative legal analysis is used to contrast traditional TDM with AI-specific processes. The article also incorporates case-law analysis and technological contextualization to assess the legal and technical compatibility of existing frameworks with AI development. This multi-method approach ensures a comprehensive understanding of both legal doctrine and technological realities.

3. Temporary acts of reproduction during AI training (InfoSoc Directive, Article 5(1))

Temporary acts of reproduction are permitted under Article 5(1) of ‘Directive 2001/29/EC’ or the InfoSoc Directive, if they are “integral and essential part of a technological process”, whose sole purpose is to enable either “a transmission in a network between third parties by an intermediary” or a “lawful use” of a work or other protected subject matter, and if they “have no independent economic significance”.

3.1 Reproduction is “temporary”, “integral and essential” and an “integral and essential part of a technological process”

The exception only justifies “temporary” reproductions that are “integral and essential”. This does not refer to the duration of the act but rather requires that the storage itself is not of long duration. The Court of Justice of the EU (CJEU) mandates that the storage “must be limited to what is necessary for the proper functioning of the relevant technological process”, and that this process must be automated in such a way that it deletes the act automatically, without human intervention, once its function is fulfilled (‘Infopaq International A/S v. Danske Dagblades Forening’, 2009). For most processes involved in training generative AI models, the requirement of merely “temporary” reproduction is not met. Web scraping and the creation and storage of a dataset are not limited to short periods. Deletion is also not automatic but depends on the will of the AI system operators. The same applies to the processing of training data during pre-training and fine-tuning; the data is not merely “temporarily” reproduced. Only the creation of copies in the computer’s RAM during reading, training and output generation can be considered temporary. Only if training occurs without creating a corpus and exclusively with data stored temporarily in RAM (with immediate and automatic deletion) can it be considered a temporary reproduction. The CJEU also requires that “the reproduction acts must be an integral and essential part of a technological process” (‘Directive 2001/29/EC’, 2001, Article 5(1)). This means, cumulatively, that the reproduction must occur entirely within the execution of a technological process and that the process could not function properly and efficiently without it. No objections are raised regarding reproductions for AI training purposes, either in preparatory acts or during training in RAM.

3.2 Reproduction serves the purpose of “lawful use” and has “no independent economic significance”

AI training does not serve the purpose of transmission in a network between third parties by an intermediary. A use is lawful if the rights holder has authorized it or if no legal restriction exists. Thus, unlike traditional TDM, AI training involves comprehensive copyright-relevant exploitation of the works used. Reproduction acts must not have “independent economic significance”. The CJEU requires that the benefit of the act must not go beyond the immediate advantage of the permitted use (‘Football Association Premier League Ltd and Others v. QC Leisure and Others’ and ‘Karen Murphy v. Media Protection Services Ltd’, 2011). In other words, the reproduction must not create a new opportunity for use. This ensures a fair balance between the author’s interest in participation and control and the interest in developing and using new technologies. For example, the German Federal Court classified the display of thumbnails in search engines (Funta, 2021; Zorluoğlu Yılmaz, 2025) as economically independent because it enabled additional revenue opportunities, especially through advertising (German Federal Court of Justice (BGH), 2017). Reproduction acts during AI training do have independent economic significance. This is evident from the high costs of collecting and preparing quality datasets and the existence of marketplaces for licensing training data (Machowicz, 2022). Moreover, the phases of training generative AI models cannot be separated without impairing the development of the desired creative-productive AI capabilities. Therefore, all training processes are indispensable and collectively fall under the concept of “independent economic significance”, from web scraping to corpus creation and storage, and through all training phases. In all the phases of training generative AI models, except possibly for operations in RAM, there is permanent reproduction of training data. The same applies to the lawfulness of use and the economic interest assessment: the various training stages cannot be separated without impairing the functionality of the developed models. Consequently, all reproduction acts associated with training have significance beyond mere use and thus constitute “independent economic significance”.

4. Text and data mining for scientific research purposes (DSM Directive, Article 3)

In the context of TDM, a distinction must be made between non-commercial scientific research and commercial mining. According to Article 3 of ‘Directive 2019/790’ (2019), or the DSM Directive, TDM for scientific research purposes is broadly privileged. This includes not only the reproduction of works but also their public accessibility. However, this exemption does not apply to commercial AI development. Activities related to commercial AI are not covered by this privilege. The same applies to the distribution of AI models that were trained using data under the exemption for non-commercial scientific research.

A recent decision by the Hamburg Regional Court in the case of ‘Kneschke v. LAION’ sheds light on the practical application of Article 3. The Court held that the reproduction of protected works during the preparatory phase of AI training, such as downloading images to verify metadata, can fall within the scope of “scientific research” under Article 3 of the DSM Directive. Notably, the Court emphasized that such preparatory acts need not lead to demonstrable research success to qualify as scientific research, provided they contribute to the broader advancement of knowledge (Manteghi, 2025).

5. Text and data mining (DSM Directive, Article 4)

The application of the exception for commercial TDM (‘Directive 2019/790’, 2019, Article 4) is not appropriate either. The frequently demanded expansion of the scope of this exception, often without a detailed examination of the technological background and context, is based on a flawed technological and conceptual assumption—namely, the blanket equation of training generative AI models with “classical” TDM. As can be demonstrated, the application of the TDM exception is incompatible with the purpose (*telos*) of the law, the wording of the Directive, the systematic structure of the legislation and the legislative history of the TDM exception. Furthermore, applying the TDM exception to the training of generative AI models would violate the three-step test under international and European copyright law (Polčák, 2020).

5.1 Key legal element: “text and data mining”

‘Text and data mining’ refers to the automated analysis of one or more digital or digitized works in order to extract information, particularly about patterns, trends and correlations (Funta and Horváth, 2022). The use of the word “particularly” implies that the scope is not limited to patterns, trends and correlations, but rather encompasses information extraction in general. The rationale behind the provision is that TDM is solely about extracting information from analyzed datasets (Peráček and Kaššaj, 2025; Kouroupis and Sotiropoulos, 2024). The justification for the exception lies in the fact that copyright protects only the creative expression of works, not the so-called semantic information contained within them. Accordingly, extracting information through TDM only affects the public domain, which is not subject to exclusive rights. Starting with an examination of the subject matter protected by copyright, it will be shown that the equation of TDM with the training of generative AI models is based on an incomplete understanding of the technological context, especially its relevance to the design of legal exceptions. The inapplicability of the TDM exception to the training of generative AI models is confirmed through a teleological and historical interpretation of the DSM Directive and its national implementations. Finally, it will be shown that the *AI Act* (‘Regulation (EU) 2024/1689’, 2024) has only limited relevance for interpreting the legal concept of ‘text and data mining’.

5.1.1 Foundation: syntax as the object of copyright protection

It is undisputed that copyright protects the creative form of expression of a work, but not the pure ideas, thoughts or theories themselves (Telec and Tůma, 2019). This principle is firmly rooted in international copyright law; it is a core component of the European copyright *acquis* and serves as a guiding principle in the jurisprudence of the CJEU (‘*SI and Brompton Bicycle Ltd v. Chedech / Get2Get*’, 2020). Our focus is on the object of copyright protection based on the specific creative form of expression, thereby distinguishing it from ideas and thoughts, as well as from technical and scientific teachings. Conceptually, this distinction between protectable and non-protectable elements is reflected in the difference between semantic and syntactic information. Semantic information refers to the meaning level, that is, the content and its significance, typically involving facts, circumstances and related statements. Syntactic information is independent of meaning, although it can carry meaning and thus include semantic content. However, as syntactic information, it only encompasses the number

of characters and their relationships to one another. Texts primarily consist of linguistic symbols (syntactic information) regardless of the meanings they convey (semantic information). Images, too, initially convey only syntactic information, such as shapes, colors and contours, regardless of the meaning of what is depicted.

5.1.2 Misconception: equating TDM with the training of generative AI models

Starting from the dichotomy between an idea and expression, as well as semantics and syntax, TDM is predominantly described in legal literature as a copyright-irrelevant use. The argument is that the non-creative information contained in a copyrighted work is not covered by the scope of copyright protection. A work is protected solely in its specific intellectual expression. Accordingly, TDM, which accesses only the semantic information of the used works, is said not to infringe on the exploitation rights granted to the author. The reproductions that precede and accompany TDM are considered permissible as necessary steps to enable TDM, which is ultimately harmless to rightsholders and desirable due to the value of the information extracted. AI training is merely a variant of TDM: since the model, even when trained on copyrighted works, only extracts “information” and does not transfer the “form of expression” in the form of syntactic information into the system, copyright law cannot prohibit such use. Doubts about whether the legislator intended to subject a fundamentally different and, at the time, largely unrecognized technology to the TDM exception are dismissed by arguing that the exception was designed to foster innovation in the private sector as openly as possible.

5.1.3 Corrections: wording, system and purpose

Extending the TDM exception to the training of generative AI models contradicts the legal framework established for the exemption of “classical” TDM. This becomes evident already in examining the wording of the law. A thorough understanding of the technological foundations (Peráček and Seehaus, 2024) also helps clarify the conceptual incompatibility of generative AI training with the system and purpose of the TDM exemption. The inapplicability of the TDM exception already follows from the wording of the law. This is often described drawing on the recitals of the DSM Directive, as the “extraction of knowledge”. TDM analyses are thus intended to extract non-copyrightable knowledge and information. However, even at first glance, it is questionable whether the training of generative AI systems,

whose internal processes are largely opaque (black box), actually leads to the disclosure of relevant information. Generative AI models are not designed for information extraction; their primary purpose is creative and productive use. In short, no one directly gains any specific information or knowledge from the training process itself. Moreover, based on the legislator's understanding of the technical processes, as reflected in the wording, the TDM exception only applies to "information" at the semantic level.

5.1.3.1 Clarification: conceptual confusion around the "right to mine"

The "mere use" of a work, referred to as "enjoyment of the work", is not subject to copyright. While this involves the use of intellectually creative content (i.e., syntax, and thus the copyright-protected substance), it is still considered legally irrelevant. This is primarily a pragmatic consequence of the fact that enforcing copyright against private users is often difficult. Moreover, pure enjoyment of a work typically occurs without reproduction, which minimizes the risk of subsequent copyright infringement. Human perception and memory are generally too incomplete and fleeting to serve as a basis for further reproduction. In contrast, the computer-based "perception" of copyrighted works during mining, understood as a "right to mine", is fundamentally different—it is questionable whether TDM is even comparable to human enjoyment of a work as a social and communicative act. The consequences are clearly different: unlike human cognition and memory, computers can store, process and reuse information completely, accurately and permanently. While this may be unproblematic in the context of classical TDM, where there is no interference with the protected substance of the work, it becomes highly relevant in the context of training generative AI models, as will be shown. Furthermore, the "right to mine" is also, in a legal sense, not equivalent to the "right to read". TDM, unlike human enjoyment of a work, does not involve the creative elements of the used works. The right to read is to be understood as the right to use even the copyright-protected, intellectually creative content expressed in the specific form of a work. It is a special and exclusive exception to the scope of protection, designed specifically for human perception and use. In contrast, TDM is defined as the use of information not protected by copyright. Therefore, both in terms of the type of content involved and the consequences of subsequent uses, human enjoyment of a work is categorically different from TDM. As a result, the authorization for one (e.g., reading) cannot in any way justify the authorization for the other (e.g., mining). This applies even more strongly to the training of generative AI models.

5.1.3.2 Misconception: apparent filtering out of syntax

Classical TDM is limited to the search and analysis of information and datasets. This process abstracts the analyzed content as semantic information, separating it from syntax, which is considered the copyright-relevant “accompanying material”. As a result, it does not infringe on the exploitation rights of authors. In practice, training of generative AI models is often described as equivalent to TDM, and thus also outside the scope of copyright law. The technical processes show that the trained AI model does not store the protected works themselves, but rather the information derived from them through machine learning. Artificial neural networks (ANNs) are said to “learn” only information about the probabilities of general characteristics of certain types of works and not acquire any information about their specific form. Therefore, it is claimed that the elements of creative expression, which fall under copyright protection, are not part of the learning process. In other jurisdictions, this discussion is framed differently in legal doctrine, but the arguments are nearly identical. In the United States, for example, a distinction is made between expressive and non-expressive use (Sag, 2019). The non-expressive use category includes, in particular, the computer-based processing of protected works. Since no use is made of the creative content, because “only the computer reads”, the copyright-protected expression is said not to be affected. This debate is framed within the fair use doctrine, which is roughly comparable to the exception provisions in European copyright law (Svoboda, 2002). The core issue is whether the rightsholder should also be granted control over the use and exploitation of the elements of their work affected by non-expressive use—those elements that go beyond mere ideas, theories and information. However, this line of reasoning mistakenly assumes that the legal distinction between idea and expression, or between information and creative content, is reflected in the technical processes and functionalities of training generative AI models. The training of generative AI models goes significantly beyond TDM in a crucial way: it captures not only semantic information, but also—and especially—creative content in the form of syntactic information, which serves as its structure. There is no filtering out of syntax, and thus no separation from semantics.

As an interim conclusion, it must be noted that the training of generative AI models clearly falls within the scope of copyright law (Rosati, 2019). In addition to the undisputed acts of reproduction involved in preparing the training process (such as web scraping and corpus creation), the actual training processes themselves are also subject to copyright law. This insight

is crucial for the application of the TDM exception. “Classical” TDM does not access the creative expression of protected works or their syntactic information and therefore does not affect the exclusive exploitation rights granted to the author. In contrast, the training of generative AI models crosses the boundary between the mere analysis of information and the extensive use of copyright-protected works. This is relevant in two key ways: At the input level, generative AI models process more than just semantic information. They also utilize syntactic elements, which are embedded into the model’s vector space. In this space, the copyright-protected components of expression and form are also represented. During the generation of creative output, generative AI models combine more than just semantic data. In particular, the syntactic structures of the training data, replicated in vector form, are recombined in the output. The resulting products are intended and capable of competing with the original works used in training and potentially displacing them from the market. This is likely why US legal scholars (Lemley and Casey, 2021) correctly distinguish between predictive and productive AI systems: for predictive systems (e.g., facial recognition, music or text recognition software), the application of the fair use defense is generally considered plausible. However, even proponents of a broad exemption (“fair learning”) doubt the legality of training productive models and systems under copyright law (Ihalainen, 2018; Noskova and Mucha, 2017).

6. Interpretation in the light of the *AI Act*

Finally, it cannot be established that the European legislator has made a binding statement in the *AI Act* (Andraško *et al.*, 2021) regarding the extension of the scope of the TDM exception under the DSM Directive. While the *AI Act* does require providers of general-purpose AI models to establish a strategy that enables compliance (Szuchy, 2024) with the opt-out rights of rightsholders under Article 4(3) of the DSM Directive, any broader, quasi “cross-instrument” statement by the legislator of the *AI Act* concerning an expansion of the TDM exception in the DSM Directive must be doubted, especially for teleological reasons. The expansive interpretation of the TDM provision in the DSM Directive by referring to a *prima facie* clarifying “interpretative aid” in the *AI Act* may seem to offer a pragmatic solution at first glance. However, this approach is questionable in light of the wording, structure and purpose of the *AI Act* (Šmejkal, 2022). First, it

should be noted that the wording of Article 53(1)(c) of the *AI Act* primarily establishes a general obligation to develop a “strategy for compliance with Union copyright law”. Only in a specific case (“in particular”) it refers to compliance with the opt-out provision in Article 4(3) of the DSM Directive. The two accompanying recitals express this in a similar way. However, a binding equivalence between TDM and the training of generative AI models cannot be conclusively derived from this. Based on the wording, it can only be concluded that providers must comply with the requirements of Article 4(3) of the DSM Directive when using TDM methods. The systematic context also reveals a distinction between TDM and the training of generative AI models. In this regard, it should be noted that the *AI Act* (Žuriš and Kohlenz, 2025) is an instrument for regulating the safety of AI applications. The legislator’s focus is accordingly on ensuring a high level of protection with regard to health, safety and fundamental rights (Vardanyan *et al.*, 2022) of the European Union including democracy, the rule of law and environmental protection. Based on this, the *AI Act* can hardly be interpreted as aiming to regulate all AI-related risks in detail (Guadamuz, 2024). Given the highly politicized legislative initiatives in copyright law in recent years it would be more than surprising if the legislator, in a procedure focused on regulating AI product safety, had simultaneously and incidentally “regulated into” the sensitive area of copyright law. The fact that the legislator was generally concerned with the rights of European authors may indeed be documented in Article 53 and the corresponding recitals. However, these normative elements can hardly be seen as aspects aligned with the regulatory objective of addressing AI-specific product safety risks. Therefore, even from a systematic perspective, the *AI Act* should not be interpreted as intended to amend or clarify the interpretation of the DSM Directive. As an interim conclusion, it must be stated that an extension of the TDM exception to the training of generative AI models can hardly be supported. Particularly significant in this regard is the fact that the legislator of the *AI Act*, as far as can be seen, did not engage in a thorough examination of the technological differences between TDM and the training of generative AI models. Therefore, the principle of the primacy of objective-teleological interpretation must remain in place.

7. Incompatibility with the three-step test

The use of copyrighted works in training generative AI models raises the question of the application of the three-step test. So far, this issue has only been discussed in preliminary stages. The prevailing view that seeks to subject the training of generative AI models to the TDM exception contradicts international and European legal requirements of the three-step test.

7.1 State of opinion

The European legislator did not foresee any remuneration obligation when designing and implementing the rules for the TDM exceptions. In literature, this is explained by the fact that TDM is considered pure information extraction outside copyright: the extracted information is not assigned to the authors, so no exception is needed, and the three-step test does not apply systematically (Zibner, 2022). Additionally, for training generative AI models, the predominant opinion holds that the application of the TDM exception is compatible with the three-step test. Only sporadically is the three-step test mentioned as a limitation on the permissibility of AI training. As far as can be discerned, the argument mainly relies on the idea that rightsholders can ensure appropriate compensation through the opt-out regulation (Medzini, 2022).

7.2 Criticism and analysis

Starting from the recognition that the training of generative AI models qualifies as a copyright-relevant act of exploitation and that the TDM exception does not apply, the current situation can be described as a copyright infringement without an exception. Therefore, an examination of the three-step test is only implied under the hypothetical assumption that the TDM exception is applicable. The training of generative AI models significantly impacts the normal exploitation of works by rightsholders. This violation of the three-step test cannot be remedied by merely granting a reservation of use or providing adequate compensation to rightsholders.

The three-step test is embedded in both international and European copyright law (Rosati, 2021). Substantially identical provisions are found in Article 13 of the *TRIPS Agreement*, Article 10 of the *WIPO Copyright Treaty*, Article 16(2) of the *WIPO Performances and Phonograms Treaty* and

Article 9(2) of the *Revised Berne Convention*. The text of Article 9(2) of the *Revised Berne Convention* states: “The legislation of the Union countries may permit the reproduction of works in certain special cases, provided that such reproduction does not impair the normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.” In European copyright law, the implementation of this test is explicitly found in Article 5(5) of the InfoSoc Directive. Article 7(2) of the DSM Directive also refers to this provision concerning the application of exceptions and limitations. This regulation adopts the three levels of the international copyright test, stating: “The exceptions and limitations mentioned in paragraphs 1, 2, 3, and 4 may only be applied in certain special cases where the normal exploitation of the work or other protected subject matter is not impaired and where the legitimate interests of the right holder are not unduly prejudiced.” Although the European Union is not a contracting party to the *Berne Convention*, the CJEU is bound by Articles 1 to 21 of the convention under Article 1(4) of the *WIPO Copyright Treaty*, to which the EU is a party (*DR and TV2 Danmark A/S v. NCB – Nordisk Copyright Bureau*, 2012).

This means that the interpretation of the scope of the three-step test and its criteria in international copyright law is also decisive for its application in European law. Exceptions or limitations to authors’ rights are therefore evaluated in three steps, based on whether (1) the limitation applies only in certain special cases; (2) the normal exploitation of the work or other protected subject matter is not impaired and (3) the legitimate interests of the rightsholder are not unduly prejudiced. These criteria must be examined cumulatively, meaning that an exemption is only valid if all three requirements are met separately. The second step of the test aims to determine whether the exception or limitation impairs the normal exploitation of the work. This question is central to the outcome of the test in all variations, particularly in the *Revised Berne Convention* and the *TRIPS Agreement*. The WTO panel decision addressed § 110(5)(B) of the *US Copyright Act* and referenced the test variant in Article 13 of the *TRIPS Agreement*. The WTO panel decision directly refers to Article 13 of the *TRIPS Agreement*. This decision represents the only direct application of the test in international copyright law to date. Although the ruling has no formal binding effect on international or national judicial bodies, it carries significant weight for interpreting the individual criteria of the test. In particular, it provides valuable insights into the definition of “normal exploitation”. Since the application of the TDM exception has already failed at the second stage, the third test step is examined only as a supplementary measure. However,

even at this level, the application of the TDM exception does not meet the requirements of the three-step test. The central question here is whether the “legitimate interests” of the rights holder are “unduly” or “unreasonably” harmed. In terms of content, the differences in wording among various international regulations are relatively minor (Funta and Plavčan, 2021). In principle, the test requires an assessment of proportionality, balancing the interests of all parties involved. Unlike the second stage, this step also allows consideration of whether the rightsholder receives material compensation.

If the TDM exception were applied to the training of generative AI models against the wording of the provision, its systematic classification and its rationale, it would result in a violation of the three-step test in international and European copyright law. The training of generative AI systems with copyrighted works impairs the normal exploitation of these works within the meaning of the second stage of the test. The normatively established opt-out cannot remedy this violation. Moreover, the application of the TDM exception also fails at the third test stage, as it results in a disproportionate infringement of the legitimate interests of rightsholders in the balancing of interests.

8. Conclusion

The legal framework surrounding copyright limitations and exceptions in the European Union presents significant challenges when applied to the training and deployment of generative AI models. The analysis has demonstrated that existing provisions, particularly those governing temporary acts of reproduction and TDM exceptions, do not sufficiently accommodate the distinct technological processes involved in AI development. Temporary reproduction requirements under the InfoSoc Directive are largely unmet due to the persistent storage of training data and the absence of automatic deletion. Additionally, reproduction acts involved in AI training bear independent economic significance, which further disqualifies them from exemption. The current state of copyright limitations only minimally accounts for the impact of training generative AI models on copyright, covering a few rare and practically irrelevant scenarios. A closer examination of generative AI technologies reveals that their exploitation of copyright-protected content in training data extends beyond semantic content processing, unlike traditional TDM.

A comparative analysis of the technological foundations of TDM and generative AI model training highlights their fundamental differences. Unlike TDM, AI training does not restrict usage to the extraction of semantic information alone. This utilization leads to the representation of training data within the model's vector space, which effectively constitutes reproduction, and thus disqualifies AI training from the TDM exception. A historical interpretation of the TDM exception reinforces this distinction. The DSM Directive's legislators did not anticipate the disruptive impact of creative AI systems, meaning that the TDM exception, originally designed for processing semantic information, cannot be extended to generative AI models that utilize syntax comprehensively. Even if one attempted to apply the TDM exception, it would still violate the three-step test in international and European copyright law (Stamatoudi and Torremans, 2021). The broad exploitation of syntactic data in copyrighted works impairs the rightsholders' "normal exploitation", constituting an irreparable violation of the test's requirements. European lawmakers must consider whether they wish to passively witness the global "race to the bottom" in copyright regulation. This is not about halting AI innovations but rather ensuring fair competition and equitable compensation for the exploited resources. As AI continues to evolve, policymakers must reassess copyright frameworks to promote fairness for rightsholders while fostering innovation. Future legal reforms should establish clear guidelines for AI training while balancing intellectual property rights (Husovec, 2016) with technological progress. Proactive adaptation will be essential in navigating ethical and commercial concerns, ensuring sustainable development in AI advancements.

Rastislav Funta is an associate professor of EU law and international law at Janko Jesensky Faculty of Law, Danubius University, Sladkovičovo, Slovak Republic.

Denisa Krásna is a PhD candidate at Bratislava University of Economics and Management, Slovak Republic.

References

- Andraško, J., Mesarčík, M. and Hamulák, O. (2021) 'The regulatory intersections between artificial intelligence, data protection and cyber security: challenges and opportunities for the EU legal framework', *AI & Society*, 36, pp. 629–636. Available at: <https://doi.org/10.1007/s00146-020-01125-5>
- Berne Convention for the protection of literary and artistic works* (1886) 9 September. Berne: WIPO.
- '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' (2001) *Official Journal* L 167, 22.6.2001, pp. 10–19. Available at: <https://eur-lex.europa.eu/eli/dir/2001/29/oj/eng> (Accessed: 3 August 2025).
- 'Directive (EU) 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' (2019) *Official Journal* L 130, 17.5.2019, pp. 92–125. Available at: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng> (Accessed: 3 August 2025).
- 'DR and TV2 Danmark A/S v. NCB – Nordisk Copyright Bureau' (2012) Judgment of the Court (Third Chamber), Case C-510/10, ECLI:EU:C:2012:244. *European Court Reports*. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62010CJ0510> (Accessed: 3 August 2025).
- Đuriš, M. and Kohlenz, J. (2025) 'AI regulatory obligations of the EU member states', *EU Law Journal*, 10(1), pp. 5–10.
- 'Football Association Premier League Ltd and Others v. QC Leisure and Others' (C-403/08) and 'Karen Murphy v. Media Protection Services Ltd' (C-429/08)' (2011) Judgment of the Court (Grand Chamber), Joined Cases C-403/08 and C-429/08, ECLI:EU:C:2011:631. Available at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:62008CJ0403> (Accessed: 3 August 2025).
- Funta, R. (2021) 'Data, their relevance to competition and search engines', *Masaryk University Journal of Law and Technology*, 15(1), pp. 119–140. Available at: <https://doi.org/10.5817/MUJLT2021-1-5>
- Funta, R. and Horváth, M. (2022) *Digital markets as a challenge for competition policy*. Brno: MSD.
- Funta, R. and Plavčan, P. (2021) 'Regulatory concepts for internet platforms', *Online Journal Modelling the New Europe*, 35, pp. 44–59. Available at: <https://doi.org/10.24193/ojmne.2021.35.03>
- German Federal Court of Justice (BGH) (2017) Judgement in Case I ZR 11/16, 21.9.2017.
- Guadamuz, A. (2024) 'The EU's Artificial Intelligence Act and copyright', *The Journal of World Intellectual Property*, 28(1), pp. 213–219. Available at: <https://doi.org/10.1111/jwip.12330>

- Husovec, M. (2016) 'Intellectual property rights and integration by conflict: the past, present and future', *Cambridge Yearbook of European Legal Studies*, 18, pp. 239–269. Available at: <https://doi.org/10.1017/cel.2016.9>
- Ihalainen, J. (2018) 'Computer creativity: artificial intelligence and copyright', *Journal of Intellectual Property Law & Practice*, 13(9), pp. 724–728. Available at: <https://doi.org/10.1093/jiplp/jpy031>
- 'Infopaq International A/S v. Danske Dagblades Forening' (2009) Judgment of the Court (Fourth Chamber), Case C-5/08, ECLI:EU:C:2009:465, *European Court Reports*. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62008CJ0005> (Accessed: 3 August 2025).
- Kouroupis, K. and Sotiropoulos, L. (2024) 'Cyber challenges amid the digital revolution in maritime transport', *Juridical Tribune – Review of Comparative and International Law*, 14(2), pp. 321–340. Available at: <https://doi.org/10.62768/TBJ/2024/14/2/09>
- Lemley, M.A. and Casey, B. (2021) 'Fair learning', *Texas Law Review*, pp. 99–181. Available at: <https://doi.org/10.2139/ssrn.3528447>
- Machowicz, K. (2022) 'The impact of the quality of information on the use of freedom of expression', *Studia Iuridica Lublinensia*, 31(3), pp. 189–201. Available at: <https://doi.org/10.17951/sil.2022.31.3.189-201>
- Manteghi, M. (2025) 'Kneschke vs. LAION: opening fresh perspectives on AI training and TDM exceptions', *European Law Blog*, 14 January. Available at: <https://doi.org/10.21428/9885764c.5f1ce2dd>
- Medzini, R. (2022) 'Enhanced self-regulation: the case of Facebook's content governance', *New Media & Society*, 24(10), pp. 2227–2251. Available at: <https://doi.org/10.1177/1461444821989352>
- Noskova, M. and Mucha, B. (2017) 'Selected issues of Slovak business environment', in *24th International Scientific Conference on Economic and Social Development – Managerial Issues in Modern Business, Economic and Social Development (ESD)*, pp. 254–259.
- Peráček, T. and Kaššaj, M. (2025) 'Legal easements as enablers of sustainable land use and infrastructure development in smart cities', *Land*, 14(4), article 681, pp. 1–39. Available at: <https://doi.org/10.3390/land14040681>
- Peráček, T. and Seehaus, R. (2024) 'The impact of the Constitutional Court ruling of 15 November 2023 on the federal government's budget planning and the potential increase in insolvency cases in Germany', *Juridical Tribune – Review of Comparative and International Law*, 14(2), pp. 178–195. Available at: <https://doi.org/10.62768/TBJ/2024/14/2/01>
- Polčák, R. (2020) *Autorský zákon: praktický komentář s judikaturou*. Praha: Leges.
- '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)' (2024) *Official Journal L* 2024/1689, 12.7.2024. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng> (Accessed: 1 August 2025).
- 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(2), pp. 198–217. Available at: <https://doi.org/10.1080/10192557.2019.1705525>
- Rosati, E. (2021) *The Routledge handbook of EU copyright law*. London: Routledge. Available at: <https://doi.org/10.4324/9781003156277>
- Sag, M. (2019) 'The new legal landscape for text mining and machine learning', *Journal of the Copyright Society of the USA*, 66, p. 291. Available at: <https://doi.org/10.2139/ssrn.3331606>
- 'SI and Brompton Bicycle Ltd v. Chedech / Get2Get' (2020) Judgement of the Court (Fifth Chamber), Case C-833/18, ECLI:EU:C:2020:461. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62018CJ0833> (Accessed> 3 August 2025).
- Šmejkal, V. (2022) *Umělá inteligence jako kvalitativní výzva pro ochranu hospodářské soutěže*. Plzeň: Aleš Čeněk.
- Stamatoudi, I. and Torremans, P. (2021) *EU copyright law: a commentary*. London: Edward Elgar. Available at: <https://doi.org/10.4337/9781786437808>
- Svoboda, P. (2002) *Vliv autorskoprávní teritoriality na vývoj evropského práva*. Praha: Karolinum.
- Szuchy, R. (2024) 'Conformity assessments under the EU AI Act: ensuring compliance, safety and trust in artificial intelligence', *Legal Business*.
- Telec, I. and Tůma, P. (2019) *Autorský zákon: komentář*. 2nd edn. Prague: C.H. Beck.
- Vardanyan, L., Stehlík, V. and Kocharyan, H. (2022) 'Digital integrity: a foundation for digital rights and the new manifestation of human dignity', *TalTech Journal of European Studies*, 12(1), pp. 159–185. Available at: <https://doi.org/10.2478/bjes-2022-0008>
- WIPO (1996) *Copyright Treaty (WCT)*, adopted 20 December 1996. Geneva: World Intellectual Property Organization.
- World Trade Organization (1994) *Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)*, 15 April. Marrakesh: WTO.
- Zibner, J. (2022) *Umělá inteligence jako technologická výzva autorskému právu*. Praha: Wolters Kluwer.
- Zorluoğlu Yilmaz, A. (2025) 'Joint controllership under the GDPR: concept, responsibilities, and liability', *Juridical Tribune – Review of Comparative and International Law*, 15(1), pp. 93–107. Available at: <https://doi.org/10.62768/TBJ/2025/15/1/06>