

## ASSESSMENT OF THE APPLICABILITY OF THE TDM EXCEPTION USING THE METHOD OF HISTORICAL INTERPRETATION

Petra Žikovská\*

**Abstract:** Submitted paper deals with text and data mining exception, which was enacted as a part of the Directive on Copyright in Digital Single Market, and its possible connection with training of generative AI systems. There have been many different opinions on the possible use of the respective directive on the training of generative AI systems among the academics and there are no decisions of supreme courts nor SDEU. Therefore, the historical interpretation method has been applied and the original will of the legislator has been investigated. During the work the analysis of a substantial amount of the written materials connected with the respective legislation has been made, including the preparatory documents as well as documents produced during the legislative process as such. Based on the thorough research and its connection with the actual historical development of the generative AI systems the conclusion has been made that the legislators in general were not aware of the possible application of the text and data mining exception to the training of generative AI systems and therefore it was not their will.

**Keywords:** historical interpretation method, copyright, Text and Data Mining Exception, generative AI systems' training

### INTRODUCTION

Artificial Intelligence (AI) in general and namely generative AI systems have the potential to change the society entirely, innovations in the field of AI continue to shape the future of humanity across nearly every industry. The legal system as such is logically delayed and the legislative instruments which would incorporate the current needs of the technological development are still being formed, following the vivid discussion among the academics and law makers. One among the recently adopted legislations is the Directive on Copyright in Digital Single Market<sup>1</sup> (DSM Directive), which was targeted to many different issues concerning the copyright-protected material in on-line environment. As such DSM Directive was not intended to cover AI issues, but as a matter of fact one of its provisions became very important for the generative AI systems, namely for the purposes of its training. This provision is one among the exceptions and limitations, which was originally included in the DSM Directive in order to help the scientific community to analyse big amounts of data through data mining. The respective exception, which is called the Text and Data Mining Exception (TDM exception), showed a very powerful potential after the DSM Directive's adoption. As a matter of fact, if this exception is applied to the AI training, it might mean that generative AI systems could be trained on the copyright-protected data free of charge, within the scope of the above-named exception. This

---

\* JUDr. et MgA. Petra Žikovská, Ph.D., Centre for Intellectual Property Law, Faculty of Law, Charles University, Prague, Czech Republic. ORCID: 0000-0003-0571-200X.

This article was written as part of the Specific University Research project of Charles University No. 260748 'Challenges of Private Law: Sustainability and Technology'.

<sup>1</sup> 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. In: *EUR-Lex* [online]. [2025-05-20]. Available at: <<https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>>.

potential effect of the TDM exception was a surprise to many academics and in connection with the implementation of the DSM Directive to the national laws, a vivid academic discussion started, with no clear answer. That is the reason, why we should start to think about the actual will of the legislators. The pivotal question is, if the legislators knew what they were voting for, namely what was the potential impact of the TDM exception in connection with the generative AI systems' training.

To start the research, let's confront the AI itself and ask it the following questions.

---

### **1<sup>st</sup> prompt: 'How is artificial intelligence trained?'**

#### **ChatGPT<sup>2</sup> answered:**

*Artificial intelligence, especially models like me (language models), is trained on a vast variety of data. This data is used to help the model understand human language, learn its structure, meanings, and contexts. Here is an overview of the main types of data typically used to train language models:*

#### **1. Publicly Available Texts**

- **Websites** – articles from Wikipedia, blogs, forums (such as Stack Overflow, Reddit), news, reviews, and technical documentation.
- **Books** – publicly available books or books with released copyright (e.g., from Project Gutenberg).
- **Scientific articles** – particularly from open databases like arXiv or PubMed.

#### **2. Public Datasets**

- *Open-source datasets created by researchers or organizations, for example:*
  - *Common Crawl* – a massive dataset from publicly crawled web data.
  - *The Pile* – a collection of text data focused on training language models.
  - *BooksCorpus, OpenWebText* – a selection of books and articles.
  - *Wikipedia* – a popular and high-quality source of factual information.

#### **3. Code and Technical Texts**

- *Repositories from GitHub and other platforms with open-source code.*
- *Documentation for programming languages and libraries.*

#### **4. Discussion Forums and Q&A Websites**

- *Stack Exchange, Reddit, Quora (if data is publicly available).*

#### **What Is Not Used (At Least Not Without Permission)**

- *Private messages, emails, and personal data.*
- *Content behind paywalls (e.g., subscription-based newspapers) or protected by copyright (such as most books and articles).*
- *Data obtained without consent or in violation of privacy principles.*

---

<sup>2</sup> ChatGPT is a chatbot developed by OpenAI, designed to generate human-like text and engage in conversational dialogue. It utilizes large language models, like GPT-4o, to understand and respond to user prompts in a natural and informative way. Essentially, it's an AI that can answer questions, write different kinds of content, and even engage in creative tasks. In: *TechTarget* [online]. [2025-05-20]. Available at: <<https://www.techtarget.com/whatis/definition/ChatGPT#:~:text=Published:%20Mar%2004%2C%202025,learning%20to%20improve%20future%20responses>>.

### **Anonymization and Ethics**

*During training, techniques are often used to anonymize or filter inappropriate content to avoid retaining personal data or harmful material. The goal is for AI to respect people's privacy and safety.*

**2<sup>nd</sup> prompt: 'I'm interested in the use of copyright-protected data in relation to the exception for text and data mining (TDM).'**

**ChatGPT answered:**

*That's a very important and timely issue – especially due to the ongoing debates around intellectual property rights versus technological advancement. Let's break it down step by step:*

#### **What is Text and Data Mining (TDM)?**

*TDM is the process where automated tools analyze large volumes of text or data files to identify patterns, relationships, meanings, or information. When training language models, TDM is often used to process large text corpora.*

#### **How This Relates to Copyright**

*Copyright-protected materials (e.g., books, articles, songs, web content) cannot be freely used for AI training – except in cases where:*

- 1. They are in the public domain (copyright has expired),*
- 2. They are under an open license (such as Creative Commons),*
- 3. There is a legal exception in place – which is the case for TDM.*

#### **TDM Exception in the European Union**

*The European Union introduced the DSM Directive (2019/790), which defines the legal framework for TDM. It contains two key exceptions:*

##### **1. Article 3 – Research and Cultural Heritage**

- Exception for research purposes.*
- Applies to universities, libraries, archives, etc.*
- Mandatory exception – cannot be contractually overridden.*

##### **2. Article 4 – Commercial Use (e.g., AI)**

- Allows any user (including companies) to carry out TDM on legally accessible content, unless explicitly prohibited (known as an opt-out).*
- Websites that do not wish to be mined must clearly indicate this (e.g., in terms of use or via robots.txt).*

#### **In Practice:**

*If content is freely accessible online and has no opt-out (i.e., no TDM prohibition), it may be used for commercial AI training. However, if access is paid or protected, it cannot be used without a license or explicit permission.*

**Summary**

<i>Question</i>	<i>Answer</i>
<i>Can copyright-protected text be used for AI training?</i>	<i>Yes, in the EU, it is allowed if the content is legally accessible and there is no TDM opt-out.</i>
<i>What if the content is paywalled or has an opt-out?</i>	<i>Then it cannot be legally used without a license.</i>
<i>Does the exception also apply to companies?</i>	<i>Yes, according to Article 4 of the DSM Directive.</i>
<i>Is the situation the same in the US?</i>	<i>No, in the US it is governed by 'fair use,' which is more complex and context-dependent.</i>

**3<sup>rd</sup> prompt: ‘I’m more interested in the impact on authors and how they can protect themselves.’**

**ChatGPT answered:**

*That’s truly a key question, because the development of technologies like AI can have a significant impact on authors, journalists, and publishers who rely on copyright protection and payment for their work. Let’s explore a few aspects of how authors can protect their works and what steps they can take if they feel harmed by the use of their copyrighted content without consent.*

**Possible Future Scenarios**

**1. Standardized Licensing Models**

*A new market for AI training licenses could emerge, where authors and publishers sell access to their texts for AI model training. This approach could benefit both sides – authors would receive fair compensation, and AI developers would gain legal access to training data.*

**2. Changes in Legal Regulations**

*If authors receive strong support and further legal disputes arise, we can expect that legal frameworks surrounding AI and copyright may evolve, possibly introducing new specific rules for AI training.*

**Conclusion**

*Authors have several ways to protect their copyrights and defend against the unauthorized use of their texts for AI training:*

- *Contractual agreements (licenses and terms of use).*
- *Legal action (cease and desist letters, lawsuits).*
- *Assertion of rights through copyright management organizations.*

*If legal frameworks become clearer and specific licensing models for AI are developed, it could lead to a win-win situation for both authors and technology companies.*

The conversation with ChatGPT was held on 16 May 2025. From the AI's perspective everything is crystal clear regarding the training of generative AI systems on copyright-protected data. The rest of the world is not as certain as ChatGPT and there are many different opinions on this issue. There are more questions than answers and the society in general is waiting for court decisions in order to get answers, which were brought by the TDM exception provisions.<sup>3</sup> As this article deals mostly with the respective exception, it should be appropriate and useful for the reader to quote it here in full text for the reference:

DSM Directive's Article 4:

*Exception or limitation for text and data mining*

1. *Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.*
2. *Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.*
3. *The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.*
4. *This Article shall not affect the application of Article 3 of this Directive.*<sup>4</sup>

This new exception, which was enacted by DSM Directive has become a pivotal centre of many academic discussions, although it caught little interest during the negotiation of the DSM Directive. The reason for such the sudden attention is the possible application of this exception during the training of generative AI systems.<sup>5</sup> In general, the generative AI systems need a large amount of a quality data in order to produce usable outputs. The exception, which is mandatory on all EU Member States has got much larger importance than anybody probably expected during the respective legislative process. The crucial question here is was it meant to be used by generative AI systems to facilitate its training? Or in another words is the exception applicable to its training? What is necessary to mention here for the quality of further analysis is that there are two types of the exception - DSM Directive's Article 3 lists the exception intended for the scientific purposes and DSM Directive's Article 4, which serves as an exception for general purposes and is subject to the possible opt-out from the side of the rightholders.

---

<sup>3</sup> DSM Directive's Articles 3 and 4.

<sup>4</sup> 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, Article 4 Exception or limitation for text and data mining. In: *EUR-Lex* [online]. [2025-05-20]. Available at: <<https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>>.

<sup>5</sup> Generative Artificial Intelligence's definition according to the Oxford Learner's Dictionaries on-line: 'the development of computer systems that can copy intelligent human behaviour and produce new content, especially text or images'. In: *Oxford Learner's Dictionaries* [online]. [2025-05-20]. Available at: <<https://www.oxfordlearners-dictionaries.com/definition/english/generative-artificial-intelligence?q=generative+artificial+intelligence>>.

## I. STATE OF PLAY

There are many answers to the above-mentioned questions, especially among academics and also there are already some court decisions. We might divide the academics' opinions into three big subgroups. The first group stipulates that the TDM exception is not applicable due to several reasons, which we will elaborate later in this paper. Other group stipulates that the exception might be hypothetically applied but does not pass the three-step test conditions and therefore it is inapplicable. And of course, there is the third group of academics, who are persuaded that the TDM exception perfectly fits for training of generative AI systems. Let's elaborate a bit on those opinions to bring broader perspective on the scope and variety of the discussions, which is rather extraordinary.

Let's start with the decision of the *Hamburg Regional Court*, which decided in the *Kneschke v. LAION* case.<sup>6</sup> The plaintiff was an author of the photograph in dispute, which was made legally available to the public on the website of "photobank", where it was placed for the further exploitation to the possible clients. The photograph was made part of the dataset collected by LAION (defendant). The datasets consisted only of hyperlinks (URL links) directed to the actual data, but during the collection process the copy of each part of the dataset was made in order to control the accuracy of the hyperlinks. The defendant was a non-profit organization, that created and made publicly available datasets for further use by the eventual commercial AI startups etc. The plaintiff, Mr. Kneschke, sought an injunction against the defendant preventing it from any further reproductions of his photo. Furthermore, the plaintiff referred to a disclaimer on the website of the "photobank" saying that any downloading, indexing, scraping, or catching of any content of the website is forbidden which the court interpreted as an opt-out.

The decision of the court was quite interesting and well commented by academics.<sup>7</sup> The court said that LAION conducts "science" as this is a non-profit organization and made the datasets available "free of charge", so the court interpreted the expression "scientific" quite broadly. Based on that, the court stated that the opt-out is not relevant, as it is not applicable to the scientific TDM exception (DSM Directive's Article 3) and might be ignored by LAION and that the behaviour of LAION was perfectly within the legal frame. Based on the above the lawsuit was dismissed. Most interesting is that the court supported its view by pointing out that AI Act's Article 53 (1)(c) and (d)<sup>8</sup> explicitly refers to the reservation of rights in DSM Directive's Article 4(3). The AI Act requires providers of general-purpose AI models to put in place a policy to comply with EU copyright and related rights rules. For the court the reference made by the European law makers was a crucial

---

<sup>6</sup> Hamburg Regional Court's decision (LG Hamburg), 310 O 227/23, issued on 27 September 2024.

<sup>7</sup> PUKAS, Jonathan, NORDEMANN, Jan, Bernd. German Regional Court of Hamburg paves the way to treat the reproduction of works as AI training data under the EU text and data mining exceptions. *Kluwer Copyright Blog*. In: *Kluwer Copyright Blog* [online]. 25. 10. 2024 [2025-05-20]. Available at: <<https://copyrightblog.kluweriplaw.com/2024/10/25/german-regional-court-landgericht-of-hamburg-paves-the-way-to-treat-the-reproduction-of-works-as-ai-training-data-under-the-eu-text-and-data-mining-exceptions/>>.

<sup>8</sup> 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). In: *EUR-Lex* [online]. [2025-05-20]. Available at: <<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>>.

sign of their willingness to use the TDM exception for AI training purposes. It seems like that the law makers felt the need to make their will clear and therefore implemented into a more recent legal instrument sentence trying to clarify the previous uncertainty. The lawmakers' will is an important part of further analysis and therefore the argument used by the judge is significant, although it is opinion of a single person only. The case has been appealed, so the decision of the higher court is awaited.

The above-mentioned court is persuaded that the TDM exception is intended for AI training purposes, on the other hand prof. Rosati<sup>9</sup> interpreted the role of the TDM exception and its possible application in the learning phase of AI training differently. According to her opinion the TDM exception does not cover the acts of reproduction and communication to the public, so if a part or a whole work of art on which the system has been trained, is also part of the output of the specific system, then the act of making available occurs, which is not covered by any exception. Regarding the exception as such prof. Rosati stressed the importance of the three-step test, which is a general interpretation rule and should be applied to all uses potentially covered by exceptions, including the TDM exception. Three-step test has been constituted as a general interpretation rule related to exceptions and its basic task is to bring *fair balance* in between different rights and interests, namely right of the authors and users of their works. The three-step test [InfoSoc Directive's Art 5(5)]<sup>10</sup> stipulates that exceptions shall apply in certain special cases only, which do not conflict with a normal exploitation of the work and it does not unreasonably prejudice the legitimate interests of the rightsholder. In the case of the training of generative AI systems on the copyright-protected materials, the amount of used copyright-protected data definitely should be taken into consideration as well as the effect on the original works and their potential to be communicated to the public. And it is in doubt if the generative AI systems which used a million units of datasets during its training would pass the three-step test conditions, which are essential for the application of any specific exception.

Interestingly some academics on the other hand stipulates that during the phase of training the traditional *use* of work of art is not conducted as the generative AI systems are looking for patterns and systems, not for the expression of the artists' mind. This is a truly interesting approach to the problem.<sup>11</sup> According to such an opinion in a properly designed copyright framework there should be no need for a TDM exception, as the extraction of "factual information" from protected content is external to the copyright.<sup>12</sup>

---

<sup>9</sup> ROSATI, Eleonora. Infringing AI: Liability for AI-generated outputs under international, EU, and UK copyright law, *European Journal of Risk Regulation*. 2024. In: *European Journal of Risk Regulation* [online]. [2025-05-20]. Available at: <<https://ssrn.com/abstract=4946312>> or at: <<http://dx.doi.org/10.2139/ssrn.4946312>>.

<sup>10</sup> 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. In: *EUR-Lex* [online]. [2025-05-20]. Available at: <<https://eur-lex.europa.eu/eli/dir/2001/29/oj/eng>>.

<sup>11</sup> MARGONI, Thomas, KRETSCHMER, Martin. A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. 2022. In: *GRUR International* [online]. Vol. 71, No. 8, pp. 685–701. [2025-05-20]. Available at: <<https://ssrn.com/abstract=3886695>> or at: <<http://dx.doi.org/10.2139/ssrn.3886695>>.

<sup>12</sup> FLYNN, Sean, GEIGER, Christophe, QUINTAIS, João Pedro, MARGONI, Thomas, SAG, Matthew, GUIBAULT, L., CARROLL, Michael W. Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action. *European Intellectual Property Review*. 2020, Vol. 42, No. 7, pp. 393–398; Research Paper No. 2020-12. In: *American University, Washington College of Law* [online]. [2025-05-20]. Available at: <<https://ssrn.com/abstract=3578819>>.

Copyright as such does protect original expressions, not the ideas, concepts, principles. So as generative AI systems in the training phase do not look for copyright-protected material as such, only for its *patterns*. But in order to learn from it there should probably be copy made and for that specific copy the exception would be needed.

To conclude these academics' opinions' overview, which is far from being exhaustive allow us to mention the study commissioned by Copyright Initiative, *Copyright Law and training of generative AI – technological and legal foundations*<sup>13</sup> with many interesting conclusions. The study was conducted as an interdisciplinary including technological and legal foundations and for the first time on such a scale a computer scientist and legal scholars jointly created evidence regarding the processing steps in AI training. This approach as such is really promising as the knowledge of the precise technological process of AI training is strongly needed in order to evaluate it from the legal perspective. This extensive study proves that training of generative AI systems is not covered by the TDM exception as *'the statutory language and text of the provision, its systematic conception, and the ratio of the exception indicate that it must not be applied to the training of generative AI models.'*<sup>14</sup>

## II. THE HISTORICAL METHOD OF LEGAL INTERPRETATION

As demonstrated above, the interpretation of the provision on the TDM exception, particularly in relation to its potential applicability to the training of generative AI systems, remains unsettled. At present, no judicial decisions have been issued that would provide authoritative guidance; academic commentaries are not able to fully fill the interpretive gap. Only recently has a Hungarian court (Referring court – *Budapest Környéki Törvényszék*) submitted a preliminary reference<sup>15</sup> to the Court of Justice of the European Union (CJEU), which, among other issues, seeks clarification on the scope of the TDM exception. The interpretive challenges surrounding this provision may also suggest that the methods typically employed by the academic community are insufficient. Most of authors addressing this issue have relied predominantly on linguistic and teleological interpretation. They have sought to ascertain the current aim and purpose of the provision in question, particularly in the context of the DSM Directive as a whole. However, the DSM Directive is substantively heterogeneous, and neither its systematic structure nor the placement of the TDM exception within it brings significant interpretive insights.

---

<sup>13</sup> DORNIS, Tim W., STOBER, Sebastian. *Copyright Law and Generative AI Training – Technological and Legal Foundations (Urheberrecht und Training generativer KI-Modelle – Technologische und juristische Grundlagen)*, Authors' Rights Initiative, 29 August 2024. Open Access book in the NOMOS Verlag (Baden-Baden) publisher's series *Recht und Digitalisierung/Digitization and the Law*. In: *Nomos eLibrary* [online]. [2025-05-20]. <<https://www.nomos-elibrary.de/10.5771/9783748949558/urheberrecht-und-training-generativer-ki-modelle?page=1>>.

<sup>14</sup> Ibid.

<sup>15</sup> Case C-250/25, *Like Company*, Request for a preliminary ruling from the Budapest Környéki Törvényszék (Hungary) lodged on 3 April 2025 – *Like Company v Google Ireland Limited*. In: *InfoCuria* [online]. [2025-05-20]. Available at: <<https://curia.europa.eu/juris/document/document.jsf?text=&docid=301042&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=6837030>>.

In the present context, it is appropriate to inquire into the actual will of the legislator and to apply the historical method of legal interpretation.<sup>16</sup> For the purposes of this analysis, the historical method is understood in its subjective variant – that is, one that does not anchor itself in the text of the law, but rather treats the statutory text as a result, and investigates its source, namely the will of the legislator. In general, every statute may be regarded as an expression of legislators' will; hence, historical interpretation seeks to uncover that will. This legislative will may be derived from the statements and conduct of those directly involved in the legislative process. Such will can be found in written materials, including explanatory memoranda, impact assessments, preparatory documents, and other legislative records. Beyond these written sources, one may also examine Parliamentary transcripts, including speeches and debates held during the legislative process. Even statements made in support of unsuccessful amendments may yield valuable insights into the will of the legislative majority. Furthermore, scholarly interviews with participants in the legislative process may also serve as a useful, albeit more subjective, source of interpretation. In such cases, it is essential to approach recollections critically, given the inherent unreliability of memory over time. Additionally, one may consider the broader circumstances surrounding the enactment of the legislation – commonly referred to as the *occasio legis*.<sup>17</sup> This involves examining the socio-political and historical context that may have influenced the legislative process. While such an approach constitutes an auxiliary element of the historical method, it cannot and should not be employed in isolation from other sources, particularly the analysis of written sources.

It must be emphasized that the historical method cannot be utilized independently of other traditional methods of interpretation – namely linguistic, systematic, and teleological analysis. Problems arise when these methods produce divergent results. Which method, then, should take precedence? At present, historical interpretation is generally categorized as a non-standard method, contrasted with the standard methods of linguistic, teleological, and systematic interpretation. However, no consensus exists – either in theory or practice – regarding the hierarchy of these interpretive tools. It is generally accepted that the conclusions reached through historical interpretation should not contradict the outcomes of standard interpretative methods.<sup>18</sup> Where there is clear evidence of legislative will, it would be inappropriate to disregard it entirely in a teleological analysis. On the contrary, findings based on historical interpretation can serve as an important corrective to teleological reasoning. For this reason, teleological arguments should, wherever possible, be substantiated by documentary evidence of legislative will. The strength and value of the historical method lie in its foundation in objectively verifiable documents that can be critically examined. Nonetheless, if a prevailing and widely accepted

---

<sup>16</sup> To more elaborate on the historical method of legal interpretation the following categories are recognised: Intentionalism is an interpretive approach that emphasizes determining the legislator's intention, trying to determine what the legislator had in mind when creating a specific legal norm. In this regard, we also distinguish between originalists, who look for the legislator's intention at the time the law was created, and counterfactualists, who ask the question of how the legislator would have decided if they had the knowledge and understanding of today's people.

<sup>17</sup> WINTR, Jan. *Metody a zásady interpretace práva. 2<sup>nd</sup> edition*. Prague: Auditorium, 2019, p. 267.

<sup>18</sup> GERLOCH, Aleš. *Teorie práva. 2<sup>nd</sup> extended edition*. Dobrá Voda u Pelhřimova: Aleš Čeněk, 2001, p. 123.

interpretation emerges over time that diverges substantially from the original legislative will – and gains broad societal acceptance – it would be a mistake to revert to the historical interpretation, particularly in light of the importance of legal certainty. It is also universally acknowledged that the older a legal provision is, the less relevant historical interpretation becomes.<sup>19</sup> As a general principle, where there is a conflict between the statutory text and legislative will, the text should prevail – again, in the interest of preserving legal certainty. For the most legal subjects, uncovering the subjective will of the lawmaker is an impossible exercise, thus, they must rely primarily on the statute’s wording.<sup>20</sup>

### III. ANALYSIS OF THE RESPECTIVE MATERIALS

#### III.1 Introduction

The purpose of this research was to answer the pivotal question if the TDM exception was also originally intended for generative AI systems’ training. In other words, if the members of the legislative bodies were aware of that possible application, more precisely what was their original will. For that purpose, the plain subjective historical method of interpretation has been used. The main sources of the analysis were written materials which were published on the official web of the European Parliament. The materials could be divided into two groups. Preparatory works’ materials and documents covering the legislative process of the DSM Directive. In the first stage the reports, results of the public consultations, committees’ newsletters, minutes from the committees’ meetings etc. have been used. The most relevant were the *Committee on Legal Affairs’* meetings, working groups’ minutes (namely *Working group on Robotics and Artificial Intelligence*), minutes from workshops, briefings requested by the different committees etc.

In the second phase the legislative procedure of the DSM Directive has been followed, namely the proposals, amendments, and also video documents from the European Parliament’s plenary meetings have been researched. There has been quite a substantial amount of the respective documents and materials to analyse, which supported the validity and relevance of the submitted analysis and conclusions. Below the most significant and relevant findings, which helped to formulate the conclusions have been picked, and are presented in a chronological order.

---

<sup>19</sup> WINTR, Jan. *Metody a zásady interpretace práva. 2<sup>nd</sup> edition*, p. 263.

<sup>20</sup> Melzer proposes the following prioritization of interpretive methods: in the case of ambiguous legal text, teleological interpretation should be preferred over formal systematic and historical approaches. Conversely, where the statutory text is unambiguous, linguistic interpretation enjoys prima facie precedence over teleological reasoning. However, this presumption may be displaced if the clearly defined purpose and rationale of the norm run counter to its plain linguistic meaning. A persuasive linguistic interpretation must yield to teleological analysis only in two circumstances: (a) where the historical interpretation offers unequivocal and explicit support for the linguistic reading and no significant change in societal conditions has occurred in the interim; and (b) where the legal order prohibits further development of the law through analogy or teleological reduction.

MELZER, Filip. *Metodologie nalézání práva. Úvod do právní argumentace. 2<sup>nd</sup> edition*. Prague: C. H. Beck, 2010, p. 233.

## III.2 Analysis of the relevant preparatory materials

### III.2.1 Working Groups' materials

On 25 September 2014, the Committee on Legal Affairs' *Working Group on Intellectual Property Rights and Copyright Reform* (WG on IPR) was set up. The aim of the WG on IPR was to reflect on intellectual property rights issues and especially to pave the way to the upcoming reform of the EU legal framework on copyright. In the minutes from meeting of the above mentioned working group (26 March 2015)<sup>21</sup> the text and data mining and its importance not only for scientific purposes, but also for a general purpose, were discussed. T. Navracsics, Commissioner for Education, Culture, Youth, & Sport, recalled that he was responsible for cultural content in terms of copyright. He said that the technological evolution in the cultural sphere called for assessing how to get easier access to cultural content and at the same time preserved the sophisticated structure of the copyright regime which remunerates creators. After this opening the need of the TDM exception was discussed. On 22 April 2015, the meeting of the WG on IPR<sup>22</sup> was held in Brussels where the exchange of views with stakeholders on copyright issues in the digital environment took place. Open access needs among academia were discussed and Ms. A. Swan, Director of Advocacy Programmes, *SPARC Europe*,<sup>23</sup> presented the issue of text and data mining in regards with the communication system of scholars and academia. Furthermore, three possible solutions were discussed (the specific TDM exception, removal of the non-commercial restriction from the scientific research exception, or the extension of temporary copies' exception) as well as the term *scientific* within the multi-linguistic society in connection with the translation machines. The WG on IPR presented the same opinion to the *Committee on Legal Affairs* on 13 June 2016.<sup>24</sup>

On 12 November 2015, WG on IPR published Overview of the European Parliamentary Research Service (EPRS) Study called *Review of the EU copyright framework – European Implementation Assessment*,<sup>25</sup> where the following argument appeared:

*The EU copyright framework is outdated in the light of subsequent technological developments: not only are the original exceptions and limitations increasingly out of line*

---

<sup>21</sup> Committee on Legal Affairs, Working Group on Intellectual Property Rights and Copyright Reform, Meeting – 26 March 2015 (from 9:00–11:00) in Brussels – Minutes. In: *European Parliament* [online]. [2025-05-20]. Available at: <<https://www.europarl.europa.eu/cmsdata/94782/minutes%20March%202015.pdf>>.

<sup>22</sup> Committee on Legal Affairs, Working Group on Intellectual Property Rights and Copyright Reform, Meeting – 22 April 2015 (from 15:00–17:00) in Brussels – Minutes. In: *European Parliament* [online]. [2025-05-20]. Available at: <<https://www.europarl.europa.eu/cmsdata/94783/april%20minutes.pdf>>.

<sup>23</sup> SPARC Europe – About us: 'SPARC Europe is one of Europe's key and long-standing voices advocating for unfettered access to research and education — for the academic and education community; for the whole of society. ... SPARC Europe is a Dutch foundation is committed to delivering on the promise of open access, open science, open scholarship and open education. ...'. In: *SPARC Europe* [online]. [2025-05-20]. Available at: <<https://sparceurope.org/who-we-are/about-us/>>.

<sup>24</sup> Working Group on IPR and copyright reform, Working Document, Copyright Reform, Presented to the Committee on Legal Affairs on 13 June 2016. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/cmsdata/104023/Working%20Document%20on%20Copyright%20\\_JURI%20meeting%2013.06.2016.pdf](https://www.europarl.europa.eu/cmsdata/104023/Working%20Document%20on%20Copyright%20_JURI%20meeting%2013.06.2016.pdf)>.

<sup>25</sup> Committee on Legal Affairs, JURI Report, 12 November 2015, p. 7. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/cmsdata/91823/JURI\\_Report\\_Nov2015v02.pdf](https://www.europarl.europa.eu/cmsdata/91823/JURI_Report_Nov2015v02.pdf)>.

*with technological developments, they are also arguably limiting the development of welfare-enhancing uses of information. Most notably, the lack of legal certainty regarding the development of Text and Data Mining represents a potentially serious gap in the EU acquis.*<sup>26</sup>

On 20 January 2015, the *Committee on Legal Affairs* decided to establish a ‘*Working Group on legal questions related to the development of Robotics and Artificial Intelligence in the European Union*’ (WG on Robotics and AI). The working group was primarily aimed at drafting civil law rules linked to this subject-matter. Its mission was to stimulate the reflection of the members on these issues by facilitating specific information providing an exchange of views with experts from many fields of academic expertise and enabling members to conduct an in-depth analysis of the challenges and prospects at stake. The input gathered by the working group was about to be put forward as a basis for future legislative activities. The last meeting was held in Brussels on 27 September 2016. The *Report with recommendations to the European Commission on Civil Law Rules on Robotics*<sup>27</sup> (Report on Robotics) was submitted to the European Parliament in January 2017 (while the Draft report was already submitted to the *Committee on Legal Affairs* on 31 May 2016)<sup>28</sup> and its introduction consists of this sentence:

*Whereas now that humankind stands on the threshold of an era when ever more sophisticated robots, bots, androids and other manifestations of artificial intelligence (“AI”) seem to be poised to unleash a new industrial revolution, which is likely to leave no stratum of society untouched, it is vitally important for the legislature to consider its legal and ethical implications and effects, without stifling Innovation.*<sup>29</sup>

Regarding the intellectual property rights, it has been said that:

*The resolution calls on the Commission to come forward with a balanced approach to intellectual property rights when applied to hardware and software standards and codes that protect innovation and at the same time foster innovation. Moreover, the elaboration of the criteria for “own intellectual creation” for copyrightable works produced by computers or robots is demanded.*<sup>30</sup>

Regarding the data and AI the Report on Robotics consists of the following: ‘*The current insufficient legal framework on data protection and ownership is of a great concern due to the expected massive flow of data arising from the use of robotics and AI.*’<sup>31</sup> The rapporteur of the above-named report MEP M. Delvaux in her interview for platform

<sup>26</sup> Committee on Legal Affairs, JURI Report, 12 November 2015, p. 8. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/cmsdata/91823/JURI\\_Report\\_Nov2015v02.pdf](https://www.europarl.europa.eu/cmsdata/91823/JURI_Report_Nov2015v02.pdf)>.

<sup>27</sup> Committee on Legal Affairs, Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103 (INL)). In: *European Parliament* [online]. 17. 1. 2017 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/A-8-2017-0005\\_EN.html](https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html)>.

<sup>28</sup> Committee on Legal Affairs, Draft Report with recommendations to the e Commission on Civil Law Rules on Robotics (2015/2103(INL)). In: *European Parliament* [online]. 31. 5. 2016 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-PR-582443\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf)>.

<sup>29</sup> Committee on Legal Affairs, Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103 (INL)), 27 January 2017. In: *European Parliament* [online]. 27. 1. 2017 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/A-8-2017-0005\\_EN.html](https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html)>.

<sup>30</sup> *Ibid.*

<sup>31</sup> *Ibid.*

Robohub (15 July 2016) said: ‘Among other that with the own-initiative report, the European Parliament aims at expressing its own position regarding the creation of a legal framework on robotics and AI by providing guidelines and recommendations to the European Commission.’<sup>32</sup> On 8 February 2017 Public consultation on robotics and artificial intelligence was launched. There all citizens and all organizations were welcomed to contribute (298 respondents participated). The consultation was closed in May 2017, and the first results were published on 13 July 2017<sup>33</sup> where 27 % of the respondents expressed their concerns about intellectual property protection on connection with robotics and AI and asked for an EU regulation in this field. Also, when searching for the opinions of the MEPs in the earlier amendments of this report one can find interesting information about the actual knowledge about AI among the members of the European parliament (MEP/MEPs). For example, MEP M. Andersson proposed the following amendment:

... recalls that in the Continental European understanding of authorship, the concept of ‘intellectual creation’ is tied to the author’s personality, meant to apply to natural persons, and therefore artificial agents such as robots and artificial intelligence shall not be considered as authors, and information produced by them shall not be eligible to copyright protection.<sup>34</sup>

### III.2.2 Committee on Legal Affairs’ Missions

On 20 May 2016, there was a mission sent to Beijing, China comprised of select members of the *Committee on Legal Affairs*, where the AI, namely the current state of the legislative instruments covering AI in China was discussed. With regard to legislation, China’s representative said that China lagged behind EU with regard to development of AI.<sup>35</sup>

A visit from the *Committee on Legal Affairs* to Silicon Valley, USA, was conducted on 30 October 2017<sup>36</sup> where Google’s presentation on ‘*The Google projects on robotics, artificial intelligence and drones*’ took place. The next meeting of this mission took place at Stanford University’s School of Law, where participants met two copyright experts, Mr. P. Goldstein, professor of law at Stanford Law School, and Mr. T. Ochoa, professor of law at Santa Clara University. The meeting started with a presentation by prof. Goldstein of the recent US legislation on copyright law focusing on the development of the notion

<sup>32</sup> WENG, Yueh-Hsuan. A European perspective on robot law: Interview with Mady Delvaux-Stehres. In: *Robohub* [online]. 15. 7. 2016 [2025-05-20]. Available at: <<https://robohub.org/a-european-perspective-on-robot-law-interview-with-mady-delvaux-stehres/>>.

<sup>33</sup> European Parliamentary Research Service, Public consultation on Robotics and Artificial Intelligence First (preliminary) results of public consultation. In: *European Parliament* [online]. 13. 7. 2017 [2025-05-20]. Available at: <<https://www.europarl.europa.eu/cmsdata/128500/juri-committee-eprs-consultation-robotics-results.pdf>>.

<sup>34</sup> Committee on Legal Affairs, Amendments 133 - 386, Draft Report, Mady Delvaux (PE582.443v01-00), with recommendations to the Commission on Civil Law Rules on Robotics. In: *European Parliament* [online]. 26. 10. 2016 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-AM-592405\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-AM-592405_EN.pdf)>.

<sup>35</sup> Committee on Legal Affairs, Mission Report following the visit of the Committee on Legal Affairs to Beijing, China, on 1620 May 2016. In: *European Parliament* [online]. 15. 7. 2016 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-CR-585809\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-CR-585809_EN.pdf)>.

<sup>36</sup> Committee on Legal Affairs, Mission Report following the visit of the Committee on Legal Affairs to Silicon Valley, USA, from 30 October to 2 November 2017. In: *European Parliament* [online]. 26. 1. 2018 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-CR-613266\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-CR-613266_EN.pdf)>.

of *fair use*. Discussion on recent case law and how US judges are adapting copyright law to the new technology followed.

On 17 September 2018, the members of the *Committee on Legal Affairs* went to Montpellier, France.<sup>37</sup> The objective of the visit was to learn (from the experience in Montpellier) how to support startups in the field of robotics and AI. The larger Montpellier has created successful incubator called “*CAP Omega*”.<sup>38</sup> The discussion with startups ranged from access to data for AI training and issues about the legal interaction of the intellectual property between researchers and startups, to services to combine communication and project management to analyse territories and electoral maps. The importance of the visit to Montpellier was evident from the *Committee on Legal Affairs’* meeting minutes,<sup>39</sup> where the members agreed to postpone the hearing on Robotics and AI until after the mission to Montpellier!

### III.2.3 Other materials

On 15 January 2015, the *Report on the implementation of Directive 2001/29/EC*<sup>40</sup> was published, where the need ‘... to enable automated analytical techniques for text and data (e.g., “text and data mining”) for all purposes, provided that permission to read the work has been acquired, still for the research and educational purposes only.’<sup>41</sup> In amendments to the above-named report many MEPs expressed their opinions on exceptions and limitations and a majority called for a revision of exceptions as there is a strong need to adapt them to a new technological environment.<sup>42</sup> The MEPs recognised that further analysis of the exceptions and limitations is necessary in order to consider how these can serve the public in the digital age.

Moreover, the general principles of the TDM exception were already part of an *Opinion of the Committee on Legal Affairs for the Committee on Industry, Research and Energy and the Committee on the Internal Market and Consumer Protection* from 7 December 2015:

... copyright exceptions and limitations should be balanced, targeted and format-neutral and should only be based on demonstrated needs, and should be without prej-

---

<sup>37</sup> Committee on Legal Affairs, Mission Report following the visit of the Committee on Legal Affairs to Montpellier, France, on 17–19 September 2018. In: *European Parliament* [online]. 12. 11. 2018 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-CR-628364\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-CR-628364_EN.pdf)>.

<sup>38</sup> The Cap Omega Incubator, Montpellier’s Tertiary Sector, Eurêka District. In: *Set Up in France* [online] [2025-05-20]. Available at: <<https://www.set-up-in-france.org/real-estate-commercial/business-and-innovation-centre-cap-omega-incubator-ict-digital>>.

<sup>39</sup> Committee on Legal Affairs, Minutes Meeting of 9 July 2018, 15.00–18.30, and 10 July 2018, 9.00–12.30 and 14.30–18.30, Brussels. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-PV-2018-07-09-1\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PV-2018-07-09-1_EN.pdf)>.

<sup>40</sup> Committee on Legal Affairs, Draft Report on the implementation of 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 (2014/2256(INI)). In: *European Parliament* [online]. 15. 1. 2015 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-PR-546580\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-546580_EN.pdf)>.

<sup>41</sup> *Ibid*, p. 6.

<sup>42</sup> Committee on Legal Affairs, Amendments 281–556, Draft Report, Julia Reda (PE546.580v02-00), on the implementation of 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. In: *European Parliament* [online]. 5. 3. 2015 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-AM-549469\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-AM-549469_EN.pdf)>.

*udice to EU cultural diversity, its financing and the fair compensation of authors; emphasises that while the use of text and data mining needs greater legal certainty to enable researchers and educational institutions to make wider use of copyright-protected material, including across borders, any EU-wide exception for text and data mining should apply only when the user has lawful access, and should be developed in consultation with all stakeholders following an evidence-based impact assessment.*<sup>43</sup>

In minutes from the *Committee on Legal Affairs'* meeting on 20 April 2016 there is information about an agenda point called '*Hearing on Robotics and Artificial Intelligence'* where experts from different areas including technology, electrical engineering, humanities and law gave speeches about robotics and AI for the *Committee on Legal Affairs'* members.<sup>44</sup> On 31 May 2016, the *Committee on Legal Affairs* published '*Draft report with recommendations to the Commission on Civil Law Rules on Robotics*<sup>45</sup> where the following sentence is expressing their concerns about AI:

*Whereas ultimately there is a possibility that within the space of a few decades AI could surpass human intellectual capacity in a manner which, if not prepared for, could pose a challenge to humanity's capacity to control its own creation and, consequently, perhaps also to its capacity to be in charge of its own destiny and to ensure the survival of the species.*<sup>46</sup>

and also:

*'The current insufficient legal framework on data protection and ownership is of great concern due to the (expected massive) flow of data arising from the use of robotics and AI.'*<sup>47</sup>

In October 2016 there was an in-depth briefing provided by the European Parliament policy department called '*Artificial Intelligence: Potential Benefits and Ethical Considerations*<sup>48</sup> on the request of *Committee on Legal Affairs*. This briefing also included a substantial definition and explanation what AI is and also its relation to works of art was mentioned.

On 22 March 2017, there was hearing organized by *Committee on Legal Affairs* with a title *Copyright reform: the new mandatory exception* where the academic perspective on

---

<sup>43</sup> Committee on Legal Affairs, Opinion for the Committee on Industry, Research and Energy and the Committee on the Internal Market and Consumer Protection on Towards a Digital Single Market (2015/2147(INI)), p. 4. In: *European Parliament* [online]. 7. 12. 2015 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-AD-557273\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-AD-557273_EN.pdf)>.

<sup>44</sup> Committee on Legal Affairs, Minutes – Meeting of 20 April 2016, 10.00–12.30 and 15.00–18.30, and 21 April 2016, 09.00–12.30, Brussels. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-PV-2016-04-20-1\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PV-2016-04-20-1_EN.pdf)>.

<sup>45</sup> Committee on Legal Affairs, Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). In: *European Parliament* [online]. 31. 5. 2016 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-PR-582443\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf)>.

<sup>46</sup> *Ibid.*, p. 4.

<sup>47</sup> *Ibid.*, p. 21.

<sup>48</sup> ROSSI, Francesca. European Parliament, Legal Affairs, Briefing, Artificial Intelligence: Potential Benefits and Ethical Considerations, October 2016. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL\\_BRI%282016%29571380\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI%282016%29571380_EN.pdf)>.

text and data mining by Mr. A. Strowel, professor at UC Louvain was given and followed by the stakeholders' perspectives.<sup>49</sup>

Other traces of the text and data mining and its potential correlations with the technological development in general one can find in *Committee on Legal Affairs'* reports and newsletters.

In a *Committee on Legal Affairs'* newsletter from February 2018 there is an announcement for a workshop on the TDM exception,<sup>50</sup> where the members of the *Committee on Legal Affairs* were informed about the technological and legal aspects of the TDM exception by prof. E. Rosati,<sup>51</sup> of Southampton University and prof. Ch. Geiger,<sup>52</sup> of Strasbourg University. Both speakers submitted extensive studies, which supported an extension of the scope of the TDM exception beyond research organizations to all those enjoying lawful access to underlying mined materials. Technological aspects of the text and data mining were discussed in detail including the different stages of mining process. Prof. E. Rosati in terms of AI said that: *'Also Artificial Intelligence and its potential have been acquiring increasing centrality. Although classical TDM and machine learning have different utility, it should not be overlooked that both use the same key algorithms to discover patterns in data.'*<sup>53</sup>

On 21 November 2018, the *Committee on Legal Affairs* gave opinion to the *Committee on Industry, Research and Energy* on the *'Proposal of Regulation of the European Parliament and of the Council establishing the Digital Europe programme for the period 2021–2027'*.<sup>54</sup> The Proposal's Article 5 was devoted to the AI and stipulated that:

*The financial intervention by the Union under Specific Objective 2. Artificial Intelligence shall pursue the following operational objectives to build up and strengthen core artificial intelligence capacities in the Union, including data resources and libraries of algorithms in compliance with data protection legislation and with the principles of security and privacy by design and by default; ...*<sup>55</sup>

<sup>49</sup> Committee on Legal Affairs, Hearing on Copyright reform: the new mandatory exceptions. In: *European Parliament* [online]. 22. 3. 2017 [2025-05-20]. Available at:

<<https://www.europarl.europa.eu/committees/cs/product/product-details/20170320CHE01161>>.

<sup>50</sup> Committee on Legal Affairs, Newsletter, Issue 46, February 2018. In: *European Parliament* [online]. [2025-05-20]. Available at: <<https://www.europarl.europa.eu/cmsdata/138766/juri-committee-newsletter.pdf>>.

<sup>51</sup> ROSATI, Eleonora. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in Digital Single Market – Technical Aspects, European Parliament, Briefing, February 2018. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL\\_BRI\(2018\)604942\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf)>.

<sup>52</sup> GEIGER, Christophe, FROSIO, Giancarlo, BULAYENKO, Oleksandr. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in Digital Single Market – Legal Aspects, In-Depth Analysis, February 2018. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL\\_IDA\(2018\)604941\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf)>.

<sup>53</sup> ROSATI, Eleonora. The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in Digital Single Market – Technical Aspects, European Parliament, Briefing, February 2018. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL\\_BRI\(2018\)604942\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf)>.

<sup>54</sup> Committee on Legal Affairs, Opinion for the Committee on Industry Research and Energy on the proposal for a regulation of the European Parliament and of the Council establishing the Digital Europe programme for the period 2021 – 2027. In: *European Parliament* [online]. 21. 11. 2018 [2025-05-20]. Available at:

<[https://www.europarl.europa.eu/RegData/commissions/juri/avis/2018/627030/JURI\\_AD\(2018\)627030\\_EN.pdf](https://www.europarl.europa.eu/RegData/commissions/juri/avis/2018/627030/JURI_AD(2018)627030_EN.pdf)>.

<sup>55</sup> *Ibid.*, p. 26.

In the *Committee on Legal Affairs' 'Opinion on a comprehensive European industrial policy on Artificial Intelligence and Robotics'* for the *Committee on Industry, Research and Energy* from December 2018 the *Committee on Legal Affairs* welcomed the European Commission's '*Communication to the Institutions on Guidance on certain aspects of Directive 2004/48/EC of the European Parliament and of the Council on the enforcement of intellectual property rights*'<sup>56</sup> but underlined the need to monitor the relevance and efficiency of rules on intellectual property rights to govern the development of AI.<sup>57</sup> The *Committee on Legal Affairs* also welcomed that the European Commission has set up the '*High-Level Expert Group on Artificial Intelligence*', comprising representatives from academia, civil society and industry to draft '*AI ethics guidelines*'<sup>58</sup> on issues such as fairness, safety, transparency, the future of work, democracy, and more broadly the impact on the application of the *Charter of Fundamental Rights*, including privacy and personal data protection, dignity, consumer protection, and non-discrimination.

### III.3 Analysis of the legislative process

It is also very important to follow the legislative process of the DSM Directive. The European Commission presented a legislative package for a modernisation of the EU Copyright rules, including a new Directive on copyright in the digital single market on 14 September 2016. The proposal consisted of the *scientific* TDM exception which enabled text and data mining carried out by research organisations for the purposes of scientific research only. Paragraph 8 of the original recitals, published as a part of the DSM Directive's proposal, consisted of a following sentence: '*New technologies enable the automated computational analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining.*'<sup>59</sup> In the Impact Assessment provided by the European Commission the text and data mining was described as: '*... the automated processing ("machine reading") of large volumes of text and data to uncover new knowledge or insights...*'<sup>60</sup> and was assessed as a powerful scientific

---

<sup>56</sup> Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee, Guidance on certain aspects of Directive 2004/48/EC of the European Parliament and of the Council on the enforcement of intellectual property rights, COM/2017/0708 final. In: *EUR-Lex* [online]. 29. 11. 2017 [2025-05-20]. Available at: <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52017DC0708>>.

<sup>57</sup> Committee on Legal Affairs, Opinion on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)). In: *European Parliament* [online]. 12. 12. 2018 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-AL-631777\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-AL-631777_EN.pdf)>.

<sup>58</sup> High-Level Expert Group on AI, Ethics Guidelines for Trustworthy Artificial Intelligence. In: *European Commission* [online]. 8. 4. 2019 [2025-05-20]. Available at: <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>.

<sup>59</sup> Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market, COM/2016/0593 final – 2016/0280 (COD). In: *EUR-Lex* [online]. 14. 9. 2016 [2025-05-20]. Available at: <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52016PC0593>>.

<sup>60</sup> Commission Staff Working Document, Impact Assessment on the modernisation of EU copyright rules, Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes. In: *European Commission* [online]. 14. 9. 2016 [2025-05-20]. Available at: <<https://digital-strategy.ec.europa.eu/en/library/impact-assessment-modernisation-eu-copyright-rules>>.

research tool to analyse large corpuses of text and data such as scientific publications or research datasets.

Stakeholders and academics were strongly divided on the proposal; public debates were held mostly on the DSM Directive's Article 17 (liability of the on-line platforms) and the DSM Directive's Article 15 (new press publishers' right). There were very few on the topic public debates regarding the TDM exception.

Nevertheless, there were some MEPs' amendments filed to the form of the TDM exception, which will be described below. Amendment No. 117 (to the recital 8) of I. Adinolfi, L. Ferrara, D. Borrelli, D. Tamburrano:

*Those technologies allow citizens, startups, researchers and journalists to process large amounts of information to gain new knowledge and discover new trends. Whilst text and data mining technologies are prevalent across the digital economy, there is widespread acknowledgment that text and data mining can also benefit citizen science, business, the research community and journalism and, in so doing, encourage innovation. However, in the Union, individuals and legal entities having authorised access to content are confronted with legal uncertainty as to the extent to which they can perform text and data mining thereof.<sup>61</sup>*

Amendment No. 119 of V. Negrescu stated: *'Text and data mining may also be carried out by any natural or legal person with lawful access.'*<sup>62</sup> In a letter from 20 December 2016 addressed to the European Commission the members of the Dutch parliamentary party asked:

*Why is the proposed exception limited to scientific research purposes and not to research purposes in a broader sense? After all, text and data mining can equally well be conducted for the research purposes of private companies (or rather their R&D departments). The above-named members of the parliamentary VVD party consider that limiting the exception to scientific research purposes alone may possibly hamper innovation since it is often initiated by private research institutes and companies as well. Is the European Commission prepared to expand the scope of the proposed exception to include, for example, both scientific and applied research purposes, regardless of whether the research is conducted by public or private institutions?<sup>63</sup>*

There was a Parliamentary plenary debate on Copyright in Digital Single Market on 11 September 2018, which was opened by a report by MEP A. Voss on behalf of the *Committee on Legal Affairs*. The debate mostly concerned the DSM Directive's Article 17 (still Article 13 at that time), but MEP M. Schaake from Netherlands added interesting point of view by saying:

<sup>61</sup> Committee on Legal Affairs, Amendments 73 – 272, Draft Report, Therese Comodini Cachia, Copyright in the Digital Single Market, p. 37. In: *European Parliament* [online]. 28. 4. 2017 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/JURI-AM-603009\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/JURI-AM-603009_EN.pdf)>.

<sup>62</sup> *Ibid.*, p. 40.

<sup>63</sup> DUTHLER, A. W. Chair of the Standing Committee for Security and Justice, Questions about the package of new EU rules on copyright. In: *European Parliament* [online]. 20. 12. 2016 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/RegData/docs\\_autres\\_institutions/parlements\\_nationaux/com/2016/0593/NL\\_CHAMBER\\_CONT1-COM\(2016\)0593\\_EN.pdf](https://www.europarl.europa.eu/RegData/docs_autres_institutions/parlements_nationaux/com/2016/0593/NL_CHAMBER_CONT1-COM(2016)0593_EN.pdf)>.

*Mr President, our copyright needs updating. But the protection of copyright, and ever more of it, is not the direction that Europe needs. This has been a depressing debate with claims of fake news flying back and forth, rock star experts and aggressive lobbying from big companies, on both sides of the argument. But this should not distract us from the substance of what is at stake. So, with 50 colleagues we seek a compromise that puts the public interest first and is proportionate between the need to enforce copyright but also to respect fundamental rights. So that means no upload filters, but proportionate remuneration for creators. It means no publishers' rights, but presumption rights. It means new and clear text and data mining exemptions, which has not been discussed much but which is crucial for the development of Europe's artificial intelligence industry, which is something of great concern as well. We also speak of a freedom of panorama and an exemption for user-generated content. So I hope for your support, so that we can move on together and show compromise can work.<sup>64</sup>*

On 12 September 2018, an amendment was adopted (in a first reading) stating that there should be possibility for the member states to adopt the TDM exception for commercial purposes under the condition of an opt-out implementation, which should be machine readable.<sup>65</sup> At this stage of the legislative process the respective exception was only voluntary so it was left to the discretion of the individual member states.

In February 2019 (after two years of negotiations) co-legislators finally agreed to a new set of copyright rules. Let us see what the opinions of the co-legislators on the TDM exception were. As it is mentioned earlier, the original European Commission's proposal consisted of three new mandatory exceptions including the TDM exception for research purposes only. The Council of the European Union (Council) asked for an optional exception (DSM Directive's Article 3a) for enabling public and private entities to use mining techniques to access data which are lawfully accessible. The European Parliament in line with the Council's position also wanted to introduce an optional exception for a public and private entities. Interinstitutional negotiations resulted in a Trialogue's agreement between co-legislators on 13 February 2019. The new DSM Directive introduced four mandatory exceptions. In addition to the three proposed by the European Commission in its proposal the co-legislators agreed to add to the EU law a general mandatory TDM exception for other purposes (DSM Directive's Article 4). The European Parliament and the Council approved the compromise text in March 2019 and in April 2019 respectively. The DSM Directive was published in the Official Journal of the EU on 17 May 2019.<sup>66</sup>

Right after the publication of the DSM Directive the European Parliamentary Research Service published '*Briefing – Copyright in the digital single market*' and in the overview one can find the following sentence:

---

<sup>64</sup> Verbatim report of proceedings, Strasbourg. In: *European Parliament* [online]. 11. 9. 2018 [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/CRE-8-2018-09-11-ITM-012\\_EN.html](https://www.europarl.europa.eu/doceo/document/CRE-8-2018-09-11-ITM-012_EN.html)>.

<sup>65</sup> Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD))(1). In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/doceo/document/TA-8-2018-0337\\_CS.html](https://www.europarl.europa.eu/doceo/document/TA-8-2018-0337_CS.html)>.

<sup>66</sup> Official Journal of the European Union, L 130/92. In: *EUR-Lex* [online]. 17. 5. 2019 [2025-05-20]. Available at: <<https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>>.

*Furthermore, in addition to the mandatory exception for text and data mining for research purposes proposed by the Commission in its proposal, the co-legislators agreed to enshrine in EU law another mandatory exception for general text and data mining (Article 4) in order to contribute to the development of data analytics and artificial intelligence.<sup>67</sup>*

## CONCLUSIONS

The submitted research relied extensively on a wide range of written materials that are publicly available on the websites of the European Union's institutions. To a certain surprise, the volume of relevant documents was considerable, and their analysis proved to be highly time-consuming. The documents were reviewed chronologically, in the order in which they were published as part of the legislative process and its preparatory phases, including related activities such as expert seminars on text and data mining and official missions abroad. Of particular significance was the work of the WG on Robotics and AI, which had been addressing issues relating to AI since 2014. It is important to note, however, that the concept of AI as understood in 2014 was closely tied to robotics, autonomous vehicles, and similar technologies. At that time, it is unlikely that the general public – including members of the European legislative bodies – could have anticipated the development of AI in the direction of generative AI systems such as ChatGPT.

For the purpose of evaluating the preceding analysis, it is also necessary to consider the historical evolution of generative AI systems and its relation to AI in general. AI in the general sense of meaning has been already used in various areas of technology for a long time and does not conflict with the intellectual property law system. On the other hand, the generative AI systems are quite new technological innovation. For example, ChatGPT was officially released to the public on 30 November 2022, although its first non-public research model was developed as early as 2018. Given that these early models were not available to the public, it would be unreasonable to expect that majority of the legislators were aware of their potential implications at that time. Other generative AI systems such as Meta's LLaMA, Google's Gemini, or the French-developed Mistral AI were only made publicly available in the first half of 2023.

On the basis of the analysis of the respective available materials, it can be concluded that legislators could and should have had a general awareness of the term *Artificial Intelligence's* meaning. The expression *should have had* is used deliberately, since AI was the subject of several expert-led seminars (e.g., *Workshop on the TDM exception*,<sup>68</sup> where the members of the *Committee on Legal Affairs* were informed about the technological and legal aspects of the TDM exception by prof. E. Rosati and prof. Ch. Geiger in February 2018), reports, and briefing materials, which MEPs were expected to get familiar with

---

<sup>67</sup> MADIEGA, Tambiama. European Parliament, EU Legislation in Progress, Briefing, Copyright in the digital single market, European Parliamentary Research Service, Members' Research Service, PE 593.564, June 2019. In: *European Parliament* [online]. [2025-05-20]. Available at: <[https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2016\)593564](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2016)593564)>.

<sup>68</sup> See footnote No. 51.

as part of their legislative duties. *Committee on Legal Affairs* frequently informed other European Parliament's committees about its findings.

The second key finding is that the link between the TDM exception and AI was already known before the DSM Directive was adopted. In this context, it is crucial to mention that AI in that sense was intended mainly to analyse huge amount of data in order to uncover new knowledge or insights. The AI was not considered to be generator of a new content as some of the contemporary AI generative models are aimed for. This premise is supported, among other proofs, by the record of a Plenary debate in the European Parliament from 11 September 2018, as well as by a comprehensive *Report on the Copyright in the Digital Single Market Directive*<sup>69</sup> published in June 2019. The latter explicitly states that the general TDM exception was introduced in order to support the development of data analysis and AI. Additional evidence of legislators' awareness of AI can be found in the minutes of a mission undertaken by part of the European Parliament's *Committee on Legal Affairs* to Montpellier, where the need for access to data for training of AI systems was explicitly discussed.

The third – and the most significant – conclusion of the analysis concerns the legislators' will with respect to the potential use of the TDM exception for training of generative AI systems. For this assessment, the historical development of generative AI systems – which generate text, images, or music in response to natural language prompts – was particularly relevant. During the preparatory phase and legislative process, such systems were not yet known to the general public. Members of the EU legislative bodies should be considered part of this public, as it cannot be reasonably expected that they possess expert-level knowledge in every area over which they legislate. It is thus conceivable that some experts were aware of the possible link between the general TDM exception and the training of generative AI systems during the legislative process. However, this knowledge cannot be presumed to extend to the legislators themselves.

Based on the above-stated findings, it is not possible to infer that the will of the EU legislators was to allow the general TDM exception to be used for the training of generative AI systems. In short, that is because such systems were not publicly known at the time of the DSM Directive's adoption or during its preparatory phase.

---

<sup>69</sup> See footnote No. 67.