HOW CAN THE OPT-OUT MECHANISM UNDER ARTICLE 4(3) OF DIRECTIVE (EU) 2019/790 (CDSMD) BE OPTIMISED TO BALANCE THE COPYRIGHT PROTECTION OF THE NEWS ARTICLES WITH THE ADVANCEMENT OF THE TEXT AND DATA MINING (TDM) ACROSS THE EUROPEAN UNION?

Law and Technology in Europe Master's Thesis

Word count: 16 204

Student: Viktoriia Voytsitska

Student number: 9477446

Supervisor: Maja Sahadžić

Second reader: Helefom Abraha

28 June 2024

Contents

Abstract……4

I: Introduction……..5

1.1 Background……..…5
1.2 Research question ……7
1.3 Research approach and methods, and thesis structure…….9
1.4 Academic relevance of research……...……..12

II: Introduction to TDM and Copyright……...……15

III: Analysis of Legal Framework……20
3.1 Analysis of the AI Act ……...……20
3.2 Analysis of Article 4(3) of Copyright Directive (EU) 2019/790……..…….21
3.3 Analysis Art. 5.3c EU DataBase Directive…..……...23
3.4 Role and Impact of Copyright in Safeguarding the Intellectual Property of News Publishers……...24

IV: Practices of implementation and exercising of Art. 4(3) of thr Directive across different EU Member States……..……29
4.1 Poland country study - Poland challenges the rule of EU copyright law…….……..30
4.2 Bulgarian country study - A Delayed but Standard Implementation.……33
4.4 Countries comparison and Possible solutions regarding the implementation of the provisions…….34

V: Legal and Practical Challenges posed by the provisions……..…...38
5.2 Legislative challenges of implementation of European regulation…….38
5.2 Practical Challenges Arising from These Provisions……..43

VI: Recommendation of Compliance strategies with respect to Ar. 4(3) of the Directive ……..48

6.1Possible steps to be considered by EU government……48

6.2 Compliance strategies for AI-companies……51

6.3 Complience strategies for news publishers and authors……..52
6.4 Solving the Challenge of Diverse Technical Protocol……53

VII.Conclusion……57

7.1 Research outcome ……..57

7.2 Final thoughts…….59

Bibliography……...61

Abstract

This thesis explores the impact of opt-out provisions under Article 4(3) of Directive (EU) 2019/790 (CDSMD) on the practice of text and data mining (TDM) of news articles. The study examines whether reserving rights (opting-out) is adequately addressed within the scope of new exceptions and how these provisions are implemented across various EU Member States. Using a combination of legal, comparative, and empirical research methods, this thesis analyzes the legal framework surrounding copyright and TDM, with a specific focus on news content.

The key findings show significant disparities in the implementation of Article 4(3) among Member States. There are notable differences between countries such as Bulgaria, which has adopted a standard implementation approach, and Poland, which has not yet implemented the directive. These differences highlight the fragmented nature of the EU's legal landscape and pose challenges for consistent TDM practices.

The research highlights the conflict between protecting intellectual property rights and promoting innovation. While opt-out provisions give news publishers the ability to manage and profit from their content, they also pose challenges for TDM practitioners, potentially impeding research and technological advancement. The thesis emphasizes the requirement for standardized technical protocols and international alignment of copyright laws to ensure a fair approach that promotes the interests of both rights holders and technological progress.

It presents policy suggestions intended to establish a copyright framework that is fair and effective, striking a balance between the concerns of news publishers and the general public's interest in accessing and using information

I. Introduction

1.1 Background

Models like GPT and Gemini are known to be trained on millions of copyrighted materials - books, images, photo and video files. This brings on a table debate around striking a balance between protecting the economic interests of copyright owners and promoting technical innovation.

Historically, copyright laws were designed to protect authors by giving them exclusive rights to reproduce, distribute, and transmit their works. Now, Text and Data Mining[1] (hereinafter - TDM) - mecchanism, which is used to train AI models by involving copying and analyzing large volumes of text and, accordingly, challenges these traditional structures.

Such rapid advances in technology require changes in copyright laws to accommodate new uses, such as TDM, while maintaining fair compensation and control for copyright owners.

The regulation of this issue varies around the world: in the USA it is "Fair use"[2], the doctrine in the EU is the concept of "exceptions and limitations" enshrined in the EU Copyright Directive[3], while others have introduced alternative regulatory regimes or are still in the process of development.

The applicability of the EU AI Act[4] in 2024 will have a big impact on how text and data mining (TDM) is used to train AI models, as it states that AI-companies may use the content for training AI-model without permission unless the rights holder has expressly waived the machine-readable format. The right to express opt-out (reserve their works from TDM activities) is stated in Article 4(3) of the Copyright Directive). This legal mechanism attempts

[1] European Parliament, In-depth analyses for the Juri Committee “The exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market- Legal Aspects”

[2] Harvard University “Copyright and Fair Use: A Guide for the Harvard Community”

[3] 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 (CDSMD)

[4] Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts

to strike a balance between the protection of intellectual property and the encouragement of innovation.

The balance between defending intellectual property rights and encouraging the free flow of information is impacted by this crucial issue. The current state of affairs could stifle creativity, restrict information access, and give the EU unfair advantages disadvantages in comparison to countries like Japan, the US, and the UK that are more open to TDM because of its legal framework[5].

In the digital age, the relationship between innovation and copyright is crucial, especially with the introduction of cutting-edge technologies like artificial intelligence (AI) and machine learning. Text and data mining (TDM) is a key component of these technologies, which enable them to process enormous volumes of data, identify important patterns, and spur innovation in a variety of industries[6]. But the addition of opt-out clauses to EU Copyright Directive Article 4(3) begs an important question: Do these clauses serve as a sufficient veto power for rightholders, or do they stifle innovative advances in European AI?

This problem is gaining considerable relevance for the authors of news articles, because the issue already affects the right to access to information[7]. Indeed, how are AI models supposed to answer questions about socio-political issues if most news writers exercise their right to opt out? Such actions will have their consequences both as content generated by the model itself, and will be reflected on society, which will consume distorted and false content.

Nonetheless, different approaches by businesses and disparities among EU Member States have resulted from the absence of official guidelines on putting these opt-out provisions into practice. Both news publishers, who want to safeguard their content, and AI-companies, who need access to large datasets for efficient AI training, face difficulties as a result of the fragmented legal landscape[8].

[5] Article19, "Balancing the Right to Freedom of Expression and Intellectual Property Protection in the Digital Age"

[6] Innovation, Intellectual Property, and Access to Knowledge by World Intellectual Property Organization (WIPO)

[7] European Commission, "Report highlights tension between intellectual property rights and scientific progress"

[8] Michael Edwards, "The Intersection of Intellectual Property Rights and Cross-Border Data Privacy" (2020)

However, businesses like Google, OpenAI, and Microsoft have created their own tools and protocols due to the lack of official guidelines on how to implement these opt-out provisions. Because of this, the process is disjointed and inconsistent, which makes it expensive and ineffective for creators to consistently choose not to participate for every entity.

By examining how Article 4(3) is applied in various EU Member States, assessing the practical ramifications for news publishers and TDM practitioners, and suggesting compliance strategies that strike a balance between copyright protection and the need for technological advancement, this study seeks to investigate these issues[9].

The aim of this thesis is to investigate the practical impact of Article 4(3) on TDM (text and data mining) practices for news articles. We will examine how the opt-out provision is implemented across EU member states to gain valuable insights into its influence on TDM activity and its potential consequences for access to accurate information[10]. The research will also explore sub-questions such as whether a dedicated opt-out right should exist for news content and how news organizations are navigating compliance strategies. Through this analysis, the research aims to provide a clearer picture of the current situation and offer potential solutions for fostering a more balanced approach that supports both copyright protection and technological advancement. Ultimately, the goal is to ensure access to a diverse and informative news landscape for European citizens[11].

1.2 Research question

The current legal framework for Text and Data Mining (TDM) of news articles raises questions about balancing copyright protection and the free flow of information[12]. Overly

[9] L. Bjur & S. Weatherall, "A tangled web: The opt-out mechanism for text and data mining in the European Union Copyright Directive", p. 183-204 (2019)

[10] A. Dimopoulos & L. Garrison, "The Challenges of Implementing the Opt-out for Text and Data Mining in the EU Copyright Directive", Journal of Intellectual Property Law & Practice (JIPLP), p. 745-762 (2019)

[11] C. Geiger, "The Opt-Out Mechanism for Text and Data Mining in the EU Copyright Directive: A Missed Opportunity?" (2018)

[12] European Commission, report "The Impact of the Text and Data Mining Exception in the EU Copyright Directive on Research and Innovation" (2021)

restrictive opt-out mechanisms could hinder innovation in AI development and limit access to accurate information, potentially leading to the dissemination of distorted content.

With all the above considered, this thesis will answer the following research question:
How can the opt-out mechanism under Article 4(3) of the EU Copyright Directive (EUCD) be optimized to balance the protection of copyright for news articles with the advancement of Text and Data Mining (TDM) practices across the European Union?

This research will investigate the following sub-questions:

(1) Is Opt-out right from TDM activities established for news and what are the consequences for AI-companies and publishers?

This question explores whether news articles what is the current regulation and if it should receive different treatment compared to other copyrighted materials under Article 4(3). Given the potential societal benefits of AI-powered analysis of news content, considering a dedicated opt-out right might be worthwhile.

(2) How are EU member states implementing the opt-out right for TDM activities related to news articles under Article 4(3) of the EU Copyright Directive[13]?

Investigating these variations will illuminate the level of consistency and potential loopholes in the current system. Understanding these differences is essential for proposing effective solutions.

(3) What compliance strategies could be adopted in response to Article 4(3) of the EU Copyright Directive, considering the opt-out mechanism and its impact on TDM practices?

Providing recommendations best on best world practices could be helpful for finding the best solution for the chellanges.

[13] 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 (CDSMD)

1.3 Research approach and methods, and thesis structure

First of all, in this thesis, I will use the **legal dogmatic research method** to analyze principles, concepts, and governing laws and to address the question of which TDM acctivities are allowed and which are prohinbitied, along with the question of wheather these are currently some special provisions regarding news matarials.

By conducting such analysis I will describe excising law governing copyright reservation of new articles from TDM activities, in particular Article 4(3) of Directive (EU) 2019/790[14], AI Act[15], Art. 5.3c EUCD[16]. Having analyzed the relationship between these principles, rules and concepts, this thesis is aiming to fill in the gap of defining if news articles will fall under the opt-out exception.

While analyzing the difficulties of Article 4(3) of the CDSMD, legal pragmatism provides a flexible, outcome-oriented paradigm that emphasizes the significance of practical impacts and societal results of legal decisions.

In this study, I will use the legal pragmatic approach to evaluate the real-world effects of Article 4(3) on Text and Data Mining (TDM) activities. I will investigate how the current opt-out mechanism may impede technological innovation and access to information, potentially conflicting with the intended objectives of copyright protection. This will help this thesis not only provide analyses of legal framework (Chapter 3), but also highligh the practical challenges (Chapter 5) and therefore will enable to provide possible solutions (Chapter 6).

Secondly, Most Different Systems Design (MDSD) **comparative method** will be used to analyse a question how Article 4 (3) was incorporated to national law across different EU Member States.

[14] 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 (CDSMD)

[15] Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts

[16] 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, art 5.3c.

The country reports will focus on contrasting ones with significantly different approaches to the same issue - Bulgaria[17], Poland[18].

Bulgaria provides a case where the implementation of the directive sparked vivid discussions within the news sector, potentially impacting TDM activity. In contrast, Poland offers a contrasting case as they haven't implemented the directive yet and appear unlikely to do so in the near future. This allows us to explore the potential consequences of implementation (as seen in Bulgaria) versus non-implementation (as seen in Poland) for TDM practices and the news landscape.

Comparison will be made with regard to following points:

(1) How implementation of Article 4(3) influenced News Publishers in the country?
This will help us understand how the opt-out provision affects the willingness of publishers to share content for TDM.
(2) If the new Directive hasn't implemented yet, how the Country is planing to regulate TDM activities?
This provides context for understanding how TDM is currently addressed in the absence of Article 4(3).
(3) In both cases: How have news publishers reacted to the current state of copyright regulation for news articles?
Understanding their perspectives can illuminate potential challenges and opportunities related to TDM.

By comparing these aspects across Bulgaria and Poland, I will receive comprehensive insights into the impact of Article 4(3) on TDM practices. This will help to find potential solutions that support both copyright protection and technological advancement in a balanced way.

Moreover, this research will also leverage existing research on potential compliance strategies proposed by various stakeholder proposed both by EU and US publishers communities and sole authors, this thesis will evaluate the findings, provide policy, recommendations and explore the reasons for unexpected similarities and differences.

[17] Ana Lazarova "The last in line: Bulgaria implements the CDSM Directive" (Kluwer Copyright Blog,2023)

[18] Paul Keller, "TDM: Poland challenges the rule of EU copyright law" (Kluwer Copyright Blog, 2024)

Particularly I will be looking at the World Wide Web Consortium (**W3C**) 'Text and Data Mining Reservation Rights Community Group'[19] and "Baseline report of policies and barriers of TDM in Europe" by FutureTDM[20]. This is done to grasp current available solutions to the problem of different technical opt-opt protocols and provide recommendation of possible comsolidated solution[21].

Analisys will be made about the following points:

(1) Declaration the reservation of TDM rights

Analyzing how do news organizations declare their opt-out status or reservation of TDM rights under Article 4(3) will reveal the mechanisms used to communicate these preferences.

(2) Expressing a TDN Policy

Examening if news organizations have a documented Text and Data Mining (TDM) policy outlining their approach to data access and permissions will sheds light on the level of formalization surrounding TDM compliance.

(3) Stakeholder's policies

How do news organizations take into account the policies of other stakeholders, such as rights collectives or individual authors, when developing their own TDM strategies is crucial for a comprehensive view of the compliance landscape.

[19] World Wide Web Consortium (**W3C**) 'Text and Data Mining Reservation Rights Community Group'

[20] "Baseline report of policies and barriers of TDM in Europe", FutureTDM, Horizon 2020

[21] Urs Galler, Silke Ernst, "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age" (working paper #2007-01 Berkman Center Research Publication)

As a result of such research, this thesis will compare the two protocols in order to achieve synergy and propose consolidated recommendations[22] on compliance strategy, that should be considered by News Publishers.

1.4 Academic relevance of research

This thesis' investigation and insights make it clear that the opt-out clauses in Article 4(3) of Directive (EU) 2019/790 (CDSMD) have a complicated and wide-ranging effect on the practice of text and data mining (TDM) of news articles throughout the European Union[23]. indications indicate that such legislative changes are unlikely to occur anytime soon. The need for a balanced strategy that upholds copyright holders' rights while encouraging innovation and the free flow of information grows more urgent as the digital landscape continues to change.

This research makes a timely and valuable contribution by addressing the pressing need for a balanced approach that protects the rights of copyright holders while promoting innovation and open access to information in the fast-changing digital environment[24]. Using a multifaceted approach that includes legal analysis, comparative studies, and empirical research, this thesis aims to bridge gaps in the current literature in several key ways:

Comprehensive Analysis of Implementation Practices - This research offers a detailed examination of how Article 4(3) is implemented across EU member states[25]. The analysis highlights the differences in national approaches and their potential impact on TDM practices. By offering a thorough analysis of implementation practices in all Member States, the difficulties news organizations encounter when navigating opt-out provisions, and the wider implications for technological advancement and harmonization of copyright laws within the EU, this thesis seeks to fill a vacuum in the current literature.

[22] Paul Keller, Zuzanna Warso. "Defining best practices for opting out of ML training" (Open Policy Brief, 2023)

[23] Urs Galler, Silke Ernst, "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age" (working paper #2007-01, Berkman Center Research Publication)

[24] Ana Lazarova, "The last in line: Bulgaria implements the CDSM Directive" (Kluwer Copyright Blog, 2023)

[25] Paul Keller, "Generative AI and copyright: Convergence of opt-outs?" (Kluwer Copyright Blog, 2023)

Challenges Faced by News Organizations - The research sheds light on the challenges news organizations encounter when dealing with the opt-out provisions. Understanding these challenges is crucial for identifying potential solutions to ensure a smoother compliance process[26].

Impact on Technological Advancement and Copyright Harmonization - This thesis delves into the broader effects of Article 4(3) on both technological advancement in TDM and the ongoing efforts to harmonize copyright laws within the EU**.** The urgency of establishing precise guidelines and standards for opting out is highlighted by the potential global standardization[27] of the EU's balanced approach found in Articles 3 and 4 of the CDSMD Directive.

Recommendations to address challenges - This research aims to provide insightful analysis and suggestions that help shape a cogent and useful framework for copyright and TDM activities using legal, comparative, and empirical methodologies[28]. The ultimate objective is to guarantee that copyright laws not only safeguard creators but also foster an atmosphere that advances science and democratizes knowledge in the digital age.

It is critical that rights holders actively manage their AI rights in this regulatory landscape, using machine- and human-readable languages to precisely define their terms[29]. AI developers are also encouraged to approach rightsholders directly for licenses, carefully navigating the complexities of copyright compliance.

This thesis emphasizes the need for a flexible legal system[30] that can keep up with the rapid pace of technological development without sacrificing the balance of interests between researchers, creators, and AI developers. The next few years will be crucial in determining the

[26]Roy Kaufman, "Protecting Commercial AI Rights is harder than you think – EU Edition" (Scholarly Kitchen, 2024)

[27] Paul Keller, Warso Z. "Defining best practices for opting out of ML training" (Open Policy Brief, 2023)

[28] Laura L. "In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?" (Spawning Blog, 2024)

[29] Bernt Hugenholts, "The New Copyright Directive: Text and Data Mining (Articles 3 and 4)",)Kluwer Copyright Blog, 2019)

[30] Urs Galler, Silke Ernst, "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age" (working paper #2007-01 Berkman Center Research Publication

direction of copyright law, technology, and creative rights management globally as the EU continues to refine its copyright framework.

In conclusion, this thesis emphasizes the critical need[31] for a legal system that can adapt to the rapid pace of technological development without compromising the balance between the interests of researchers, creators, and AI developers. As the EU refines its copyright framework, the next few years will be crucial in shaping the future direction of copyright law, technology, and creative rights management on a global scale.

[31] Peter Mezei, "A saviour or a dead end? Reservation of rights in the age of generative AI" (Social Science Research Network, 2023)

II. Introduction to TDM and Copyright

In today's world data is the most valuable resource. To place things in context, according to an IBM marketing study, 90 percent of the data in the world today has been created in the last few years alone. Every day, 2.5 quintillion bytes of data are created, and it is expected that such growth rate will continue at an even faster pace in the future[32].

Considering the nuber of excisting data, the value of data no longer lies in data or text considered in their isolation, but rather in the extraction of such value[33]. This requires that text and data be analysed, to thus enable the discovery of new patterns and relations. Such task would be virtually impossible to perform manually, and that is where Text and Data Mining (TDM) comes into consideration.

Text and Data Mining (hereinafter - **TDM**) is a sophisticated analytical process used for training AI-model that involves copying large amount of data, extracting the relevant data and combining it to identify patterns[34].

The variety of TDM techniques, practices and end-goals makes it virtually impossible to provide a general and exhaustive illustration of how TDM works. By means of a necessary simplification, it appears however possible to distinguish three common – yet not all necessary – steps in TDM processes[35]:

STEP 1 - Access to content

STEP 2 - Extraction and/or copying of content

STEP 3 - Mining of text and/or data and knowledge discovery

---

[32] IBM Marketing Cloud (2017), «10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations».

[33] IDC (2014), «The Digital Universe»

[34] Bernt Hugenholts P., "The New Copyright Directive: Text and Data Mining (Articles 3 and 4)" (Kluwer Copyright Blog, 2019)

[35] European Parliament, in-depth analysis "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects"

As for an example IBM Watson Explore used TDM technics to perform variety of tasks like: Improve productivity in the workplace, Increase efficiency in public health management, e.g. in Italy or even Predict (correctly) who would win Italy's best-known singing competition[36]. Added value of TDM technics can not be denied, as it analyzes amount of data that human being would never be capable of.

One one hand, TDM is particularly significant in news journalism as not only enhances journalistic research but also powers content recommendation engines, helps in sentiment analysis, and facilitates the automated summarization of news articles, making the overwhelming amounts of daily news more accessible and analyzable[37].

At the same time, let's not forget about added value for society – it's undeniable that receiving information has been make easier when AI-models entered the market[38]. Now with just few sentences you can receive quite an open answer to any question quicker and better than using simple Google search. With this reason in mind, providing relevant and correct information is critical, meaning that AI-models should have access to the data to train and update the knowledge base.

One the other hand, it is known fact that companies like ChatGPT and Gemini use TMD to train their model on millions of **copyrighted** materials-books, images, created by artists, writers, photographers and journalists to make capable of solving variety of tasks. such activities can pose significant risks with respect to copyright rights of both publishers and authors.

The legal challenges of training Large Language Models (hereinafter - **LLMs**) lies directly in using TDM or so called "web scraping", as it was highlighted in discussions around the legality and ethical implications in Europe[39]. These practices face scrutiny due to potential violations

[36] Bernt Hugenholts P., "The New Copyright Directive: Text and Data Mining (Articles 3 and 4)" (Kluwer Copyright Blog, 2019)

[37] University of Turku (2023) 'Text and data mining (TDM) in the EU: What you need to know about copyright law and data analysis'

[38] Paul Keller, Zuzanna Warso. "Defining best practices for opting out of ML training" (Open Policy Brief, 2023)

[39] University of Turku (2023) 'Text and data mining (TDM) in the EU: What you need to know about copyright law and data analysis'

of the sui generis database rights established by the Data Base Directive EU and uncertainly around new opt-out mechanism presented in EU Copyright Directive.

Historically, the expression of ideas has been protected by copyright law, but not the ideas themselves or accurate information. TDM procedures. It protects works from unauthorized use, including[40]:

- **Reproduction -** The right to copy the work in any format (e.g., digital, print).
- **Distribution -** The right to make copies of the work available to the public.
- **Public Communication -** The right to make the work accessible to the public through various means (e.g., online,in presentations).
- **Adaptation -** The right to modify the work by creating derivative works (e.g., translations, summaries).

However, these laws also allow for certain uses that are essential to democracy, such as criticism, commentary, and reporting[41] as well as research, and education purposes.

- Journalists and citizens have the freedom to analyze and critique news articles, which encourages open debate and accountability. For example, a news outlet might publish an editorial criticizing the factual accuracy or bias present in another news article.
- News organizations can enhance existing content by incorporating snippets of news articles or factual information to create new reports. This allows for a more comprehensive picture of current events. Imagine a news report about a political debate that includes quotes and summaries from different news articles covering the event.
- Researchers and educators can use news articles for TDM activities to uncover trends, analyze public discourse, and inform research findings[42]. For instance, researchers might analyze news coverage of a specific policy issue to understand public opinion and its evolution over time.

[40] Paul Keller, "Generative AI and copyright: Convergence of opt-outs?" (Kluwer Copyright Blog, 2023)

[41] The Copyright Alliance (n.d.) 'Copyright and Journalists'

[42] Laura L. "In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?", Spawning Blog (Feb 2024)

The intersection of copyright law and TDM deals with the legal permissions needed to use copyrighted materials like news articles, broadcasts, and digital media for data mining purposes[43]. Copyright law safeguards the intellectual property rights of creators, including journalists and publishers, by giving them exclusive rights to utilize, reproduce, distribute, and display their works[44] [45]. However, these exclusive rights are challenged by the needs of researchers and technologists who want to use these works in ways not initially intended by copyright laws.

In most jurisdictions, news articles are protected under copyright from the moment they are created, as long as they meet the originality criterion. This means that they must be independently created and exhibit a minimum degree of creativity. The protection provided by copyright typically extends to the expression of ideas, facts, and data within the articles, rather than the facts themselves. The facts remain in the public domain.

The theoretical basis for copyright law concerning TDM depends on the concept of "fair use" in some jurisdictions and "exceptions and imitations" in others, such as those provided in the EU Copyright Directive[46]. The law aims to prevent the stifling of technological innovation that could result if access to copyrighted materials were overly restricted, thereby supporting both economic growth and the public interest in broad access to information.

Interium conclution

The explosion of data has fueled Text and Data Mining (TDM), a powerful tool for extracting knowledge. While TDM benefits fields like news journalism, its use in training Large Language Models (LLMs) raises copyright concerns. LLMs rely on copyrighted materials, potentially conflicting with creators' exclusive rights.

[43] Laura L. “In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?”(Spawning Blog, 2024)

[44] European Union. 2019. 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 2001/61/EC and 2009/28/EC Official Journal of the European Union L 117 (17.4.2019): 1-78.])

[45] World Intellectual Property Organization. 2023. "What is copyright?" Accessed June 18, 2024

[46] Mandatory,” in this case, applies to EU member states, meaning these or similar provisions must eventually be enacted by all EU member states and by future EU member states.

Nowadays, with all the technical developments deploying on everyday basis, copyright law should experience quite a major reframing to uphold to the changes. Attempts to do that can be further seen bellow in the next chapter.

III. Analysis of Legal Framework

The chapter will explore the legal background and theoretical frameworks underpinning TDM, with a special focus on news articles. This includes a detailed look at the copyright laws governing news content and the rationale behind the opt-out provisions in Article 4(3), addressing the necessity of such rights for news publishers. Thesis focus on news articles, outlining how subsequent chapters will explore the complex interplay between TDM activities and copyright law's opt-out provisions, specifically impacting news content.

3.1 Analysis of the AI Act

The EU Parliament has adopted the Regulation (EU) 2021/2068 of the European Parliament and of the Council of 20 October 2021 on Harmonised Rules for Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, European Parliament and Council of the European Union (hereinafter - **Al Act**).

According to the AI Act, providers of generative artificial intelligence models such as large language models (LLMs) can use content for learning purposes without obtaining permission, unless the rights holder has expressly waived the machine-readable format (e.g. , excluding robots.txt)

At the same time, several additional requirements related to copyright protection are established[47]:

- Compliance with the Copyright Directive (EU) 2019/790, in particular to ensure reservation of rights within the framework of the refusal mechanism provided for in Article 4(3).
- AI developers are required to develop and publish a detailed content statement, which is used to train their general purpose AI models. The transparency obligations outlined may help rightsholders protect their rights in the digital age[48].

[47] Mezei P., "A saviour or a dead end? Reservation of rights in the age of generative AI", Social Science Research Network, p.1-13 (February 2023)

[48] Laura L. "In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?" (Spawning Blog, 2024)

- The AI Office will be given the role of monitoring compliance with EU copyright law obligations and publishing summary data on teaching.

3.2 Analysis of Article 4(3) of Copyright Directive (EU) 2019/790

New provisions have been introduced by the EU Copyright Directive, namely the Digital Single Market Directive (Directive (EU) 2019/790) (hereinafter – **The Copyright Directive**), introducing two main provisions - exceptions to copyright (Article 3) and the right of refusal (Article 4).

The provisions of Article 3 of the Directive establish exceptions to the main law on copyright, allowing the use of copyrighted materials in the TDM process by research organizations and cultural heritage institutions for scientific and research purposes[49].

At the same time, Article 3 allows TDM only with respect to works or other objects (for example, databases) to which the beneficiary organizations "have legal access". According to clause 14, "lawful access" covers access to content under contractual agreements (such as subscriptions or open access licenses) as well as "content that is freely available online".

Copyright holders have the option to refuse permission to use their works for TDM purposes, as set out in Article 4(3) of the EU Directive[50].

The opt-out mechanism is designed to promote innovation and accessibility throughout the digital single market, while also supporting the broader policy goals of the European Union utlined in the EU Strategy for a Digital Single Market[51], which emphasizes fostering innovation

[49] Ana Lazarova, “The last in line: Bulgaria implements the CDSM Directive” (Kluwer Copyright Blog, 2023

[50] The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (2018) by Paul Keller et al., European Parliament, Directorate General for Internal Policies, Policy Department A: Citizens' Rights and Constitutional Affairs

[51] Urs G., Silke E., “EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age”, working paper #2007-01 Berkman Center Research Publication

and technological advancement as key drivers of economic growth and societal well-being[52]. With this approach, copyright holders can protect their economic interests, while technological advancements and research that benefit the public and the innovation landscape are not stifled[53].

The need to protect property rights and control the use of works by copyright owners justifies the inclusion of disclaimer clauses. Thus, this opt-out mechanism recognizes the property interests of copyright owners. It is the implementation of the "neighboring right" or "link tax" for rights holders that is a key component that guarantees them payment for the digital use of their publications. To clearly communicate these limitations to potential users in a digital environment, opt-out is typically implemented by marking the content in a machine-readable way, such as through metadata[54].

Such provisions impact the use of copyrighted materials, especially news content, in digital environments. The implementation of a "neighboring right" or "link tax"[55] for press publishers is a crucial component that guarantees them payment for the digital use of their publications.

Copyright holders have the option to refuse allowing their works to be used for TDM purposes, with the exception of scientific research, as stated in Article 4(3) of the Directive.[56] This clause gives authors the option to specifically bar their creations from TDM activities—that is, activities that don't involve scientific research. In an effort to strike a balance between the more

[52] European Commission. 2015. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A Digital Single Market for Europe. COM(2015) 192 final. Brussels, 6.5.2015.

[53] European Parliament, Directorate General for Internal Policies (2018) 'The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects' Accessed 19 May 2024]

[54] Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732

[55] "Appeal for action on violations of the Berne Convention by the application to copying of creative works for AI development of the TDM exception in Articles 3 and 4 of the 2019 EU Directive on Copyright" by Authors Coalition (July 2023)

[56] The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (2018) by Paul Keller et al., European Parliament, Directorate General for Internal Policies, Policy Department A: Citizens' Rights and Constitutional Affairs

general objectives of innovation and information access, this opt-out mechanism recognizes the proprietary interests of copyright holders. To clearly communicate these restrictions to potential users in the digital environment, the opt-out is usually implemented by marking the content in a machine-readable way, such as through metadata[57].

The need to safeguard the property rights and control over use of works by copyright owners justifies the inclusion of opt-out clauses[58]. These provisions act as a check on the broad exceptions made for TDM activities, making sure that the widespread and automated use of creators' works in ways that might potentially compromise established copyright safeguards does not unduly jeopardize their financial interests.

3.3 Analysis Art. 5.3c EU DataBase Directive

The European Union Database Directive (EU Directive 96/9/EC), in addition to copyright law, is a major factor in the protection of news content. This Directive is primarily concerned with protecting database investments, such as those made by news aggregators that gather and arrange enormous volumes of news content. The Directive enhances copyright law in a number of significant ways[59].

While copyright **protects the original expression** within a news article, the Database Directive goes a step further by protecting the selection and arrangement of a substantial collection of news articles[60]. TDM may involve the reproduction, translation, adaptation, arrangement, and any other modification of a copyrighted database, which means the original selection and arrangement of the contents of the database[61].

[57] Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732

[58] Ana Lazarova, "The last in line: Bulgaria implements the CDSM Directive" (Kluwer Copyright Blog, 2023)

[59] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

[60] The Copyright Alliance, "Copyright and Journalists", [Accessed 19 May 2024]

[61] The Database Directive and News Aggregation: Striking a Balance Between Rights and Innovation (2013) by Martin Kilian et al., European Journal of Law and Information Technology, Vol. 6, No. 1, pp. 1-22

This makes sure that news aggregators' time, labor, and financial resources that they used to gather and arrange these articles are acknowledged and safeguarded. This protects the database as a distinct entity and covers the particular way that the articles are chosen and organized.

The Database Directive **prohibits the unauthorized extraction** or re-utilization of substantial parts of a database, including those containing news articles. Thus, TDM may infringe sui generis database rights, including the extraction and reuse of substantial portions of the database.

In this context, even if extraction occurs without the reproduction of the original materials, the extraction itself violates the exclusive rights granted to the owner of the database. This position was expressed by the Court of the European Union, stating that the transfer of data from one medium to another and their integration into a new medium is an act of removal.

This stops unapproved third parties from simply duplicating and sharing the results of a news aggregator's labor. By doing this, it preserves the integrity of database creators' collections and safeguards their financial interests[62].

Lastly, the Directive gives the right to negotiate licensing agreements with other entities wishing to access their compiled content, including the ability to control who can extract and re-utilize their databases[63].

3.4 Role and Impact of Copyright and opt-out provisions in Safeguarding the Intellectual Property of News Publishers

As copyright gives news publishers the legal means to manage and charge for their content, it is essential to protecting their intellectual property[64]. Law helps to ensure that

[62] Paul Keller, Warso Z. “Defining best practices for opting out of ML training” (Open Policy Brief, 2023)

[63] The Database Directive and News Aggregation: Striking a Balance Between Rights and Innovation (2013) by Martin Kilian et al., European Journal of Law and Information Technology, Vol. 6, No. 1, pp. 1-22

[64] Kristen G., et al “Training AI models on synthetic data: no silver bullet for infringement risk in the context of training AI systems (Part 3 of 4)”, Claeary IP and Technology Insights (Jan 2024)

publishers can make money from their work, which is essential to the sustainability of the journalism industry, by granting exclusive rights to reproduce and distribute news articles[65].

Additionally, the protection provided by copyright promotes investment in journalism, assisting in the production of varied, excellent content[66]. Additionally, it gives publishers the authority to bargain for the terms under which their articles may be used, especially in the digital sphere where content can be shared quickly and extensively. In the era of digital distribution and reproduction, copyright enforcement gives publishers the ability to take legal action against unauthorized uses.

The Directive (EU) 2019/790 (CDSMD)'s opt-out clauses in Article 4(3) have a big impact on how text and data mining (TDM) is used in news journalism.

These clauses affect the breadth and depth of TDM operations by enabling copyright holders, such as news publishers, to prevent their works from being used in TDM procedures[67]. Journalists can safeguard their intellectual property rights and ensure that their work is not misused without due credit by choosing to opt out of having their articles used without permission.[68].

Practically speaking, the opt-out clauses have forced many news publishers to reevaluate their business plans, especially with regard to licensing and access to their digital archives[69]. Some publishers have started to offer TDM-specific licenses or subscriptions, which allow restricted data mining under regulated conditions, in order to comply with these provisions. By generating new revenue streams and upholding copyright, this strategy enables publishers to monetize their content while still having control over how it is used.

---

[65] Kaufman R., "Protecting Commercial AI Rights is harder than you think – EU Edition", Scholarly Kitchen

[66] Paul J. Heald, opyright and the Financing of Journalism ([invalid URL copyright and the financing of journalism, Oxford University Press (2016)

[67] Rochelle C. Dreyfuss, Copyright and Innovation: The Struggle for Balance ([invalid URL copyright and innovation the struggle for balance ON Oxford University Press (2017)

[68] Lih-Fen Lin et al., Copyright in a Digital Age, Oxford University Press (2018)

[69] The Impact of Text and Data Mining on the News Industry: New Business Models and Challenges (2021) by COMMUNIA Association ([invalid URL copyright text and data mining ON COMMUNIA communia-association.org])

This control is essential to preserving the financial worth of their contributions and preventing the exploitation of their creative work in ways that might jeopardize their financial and professional interests. Furthermore, by preventing the misuse of content that can result in false information or the decontextualization of journalistic work, the opt-out feature upholds the ethical standards of journalism[70]. This mechanism protects individual content creators' rights within the digital ecosystem, which in turn supports journalism's sustainability and integrity.

At the same time the restriction on data availability is one of the main effects of these opt-out clauses. The data pool that is available for TDM is limited by publishers' opt-out options, which may distort research findings or technological advancements that depend on large-scale data sets[71]. This restriction is especially troublesome in domains like machine learning, where robust and large-scale datasets are essential for training precise and reliable models. Thus, a major obstacle to the development of these technologies is the limitation on data availability.

Moreover, the opt-out clauses also draw attention to the inherent conflict between advancing innovation in data-driven technologies and defending news publishers' intellectual property rights[72]. These clauses give publishers control over how their content is used, protecting their financial interests in the process, but they can also stifle innovation by limiting access to information that is critical to the development of new technologies. It is difficult to strike a balance between these conflicting interests because, in the digital age, innovation must be encouraged while also protecting intellectual property[73].

[70] Poynter Institute, Journalism Ethics [Accessed 19 May 2024]

[71] Mezei P., "A saviour or a dead end? Reservation of rights in the age of generative AI", Social Science Research Network, p.1-13

[72] Yannis Manolopoulos et al., Text and Data Mining for the Social Sciences: Big Data Applications in Research, Edward Elgar Publishing (2017)

[73] Urs G., Silke E., "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age", working paper #2007-01 Berkman Center Research Publication

Interim conclusion

This chapter's discussions highlight the need for precise, standardized procedures and a thoughtful approach to copyright exceptions in order to guarantee that the digital environment promotes the development of data-driven technologies as well as the preservation of journalistic content.

Encouraging innovation and information access, the legal frameworks created by the EU Legal framework seeks to safeguard the intellectual property rights of publishers and journalists. Analyses of Article 4(3) of the EU Copyright Directive clarifies that **news articles formally fall within the exceptions** for Text Data Mining (TDM) activities.

The article is more than just a legal provision, it represents a significant change in how copyright law deals with new technologies and data-driven practices. The EU recognizes both the proprietary rights of content creators and the societal benefits of TDM by permitting an opt-out[74]. As TDM progresses with advancements in AI and machine learning, the implementation of this article will probably play a crucial role in shaping the digital copyright management landscape, influencing how news content is accessed and utilized in the digital age[75]. The implications of this provision are extensive, reaching beyond the EU's economic, technological, and legal domains.

The AI Act's wider implications and Article 4(3)'s opt-out provisions underscore the ongoing conflict between promoting technological advancements and defending the proprietary rights of news content creators[76]. The EU Data Base Directive goes even further in the ensuring that publishers will receive a profit from TDM activies, which is critical for sustainable journalism.

All things considered, the opt-out clauses in Article 4(3) raise issues and concerns about the use of TDM in news reporting going forward. These provisions highlight the difficulties in policing copyrighted content use in the context of quickly developing technologies by

[74] Gillespie, Marie-Therese. 2023. "Text and Data Mining in the Digital Age: Copyright Challenges and Opportunities in the EU." European Journal of Law and Technology 14 (2): 123-141.])

[75] Mercatus Center at George Mason University. 2022. "Text and Data Mining and the Future of Copyright." Mercatus Working Paper.

[76] Laura L. “In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?”, Spawning Blog (Feb 2024)

restricting data availability and creating a conflict between protection and innovation[77]. Therefore, it is essential to carefully weigh these implications in order to make sure that policies strike a suitable balance between upholding rights and promoting progress.[78]

[77] Arvind Narayanan, Bias in Big Data, Springer (2017)

[78] D. W. Milligan, Newspapers in a Digital Age, Cambridge University Press (2017)

IV. Practices of implementation of Art 4(3) across different EU Member States about news acrticles

Understanding how copyright law affects Text and Data Mining (TDM) activities, particularly with regard to news articles, requires an understanding of how Article 4(3) of the EU Copyright Directive (Directive (EU) 2019/790) is implemented in the various EU Member States. This chapter looks at how these provisions must be put into practice, offers country studies from particular Member States, and discusses the real-world difficulties and solutions in putting these regulations into effect.

Every Member State is obliged to synchronize its domestic copyright laws with the mandates of the Directive. This involves aligning their legal frameworks with Article 4(3) by incorporating the TDM exception and the opt-out provision. The way the provision is implemented should guarantee that it is understandable and enforceable in the context of the country[79].

Although Article 4(3) is implemented separately by each Member State, more harmonization is required to avoid a disjointed legal environment throughout the EU[80]. It is recommended that member states coordinate their strategies and implement best practices to guarantee uniformity and legal clarity for transnational TDM practitioners.

Moreover, it is necessary for Member States to set up a system that allows copyright holders to properly utilize their opt-out right[81]. Publishers should be able to indicate their wish to have their works excluded from TDM activities through an easy-to-use mechanism. Although there are some variations in the specifics of how this opt-out mechanism is put into

[79] Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732

[80] The Challenges of Copyright Law Harmonization in the Digital Single Market (2018) by Eleftheria Marina Moschidou,European Intellectual Property Review, Vol. 50, No. 12, pp. 837-847

[81] The Text and Data Mining Exception in the EU Copyright Directive: Best Practices for Implementation (2020) by COMMUNIA Association

practice, it usually entails notifying the public or pertinent authorities through a formal declaration or registry[82].

Robust mechanisms for monitoring and enforcement are necessary for effective implementation. Member states are required to guarantee that copyright holders' opt-out rights are upheld and that infractions are dealt with properly. This entails setting fines for noncompliance and offering rightsholders who have their opt-out notices disregarded redress.

4.1 Poland country study - **Poland challenges the rule of EU copyright law**

State of incorporation of the new provisions

More than 2.5 years after the implementation deadline, Poland is the only Member State not to implement the provisions of the Copyright Directive[83] into national law.

What is especially stunning in the Polish implementation proposal is that it not only excludes the creation of AI models from the scope of the Article 4 exception (exception for commercial uses) but also from the scope of the Article 3 exception (exception for research uses) related to TDM activities which seem especially short-sighted. This exclusion goes beyond simply limiting access for commercial AI developers. By also excluding non-profit scientific research, it raises concerns about stifling advancements in AI research that could benefit society as a whole.

Several potential explanations for Poland's stance warrant further investigation. While government claims suggest concerns about the potential misuse of AI for copyright infringement a more nuanced understanding is necessary[84]. Here, academic literature can provide valuable insights. For instance, some scholars argue that the Polish government might

[82] Bertuzzi L. "AI Act: EU Commission attempts to revive tiered approach shifting to General Purpose AI", EURACTIV

[83] European Union. 2019. 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 2001/61/EC and 2009/28/EC

[84] Polish Ministry of Culture and National Heritage, 2023

be apprehensive about the potential economic implications of unrestricted TDM activities for news publishers[85].

The government itself claims that the delay allows you to propose a better implementation[86] - to properly consider the impact of generative AI on copyright and come to the conclusion that training generative AI systems on copyrighted works does not in fact fall within the scope of the text and data mining exceptions contained in the directive.

Moreover, they claim that the capabilities of AI have increased from the initial proposal of the directive in 2019, meaning - when [*"works" with artistic and commercial value comparable to real works, i.e., man-made, are beginning to be created with the help of this technology, it seems fair to assume that this type of permitted use was not conceived for artificial intelligence. An explicit clarification is therefore introduced that the reproduction of works for text and data mining cannot be used to create generative models of artificial intelligence.]*[87]

Reaction

The possible reason for such actions could be found in Alek Tarkowski's LinkedIn post which claims that this approach will make Poland the market leader. It is also stated that Poland is working on #PLLUM project, creating a Polish language model, so if these regulations are passed, the Polish PLLUM model will speak in archaisms, after training in the public domain. From this, we can see, that such actions of the government are highly supported by AI companies and are even believed to provide huge benefits.

At the same time, the economic consequences of Poland's approach could be detrimental in two ways: Reduced Competitiveness of the Polish AI Industry and Loss of Potential Revenue Streams.

[85] Górski, Jan M. 2023. "The Copyright Directive and Text and Data Mining in Poland: Balancing Interests in the Digital Age." Journal of Intellectual Property Law & Practice 18 (6): 489-502

[86] Twardowski, Piotr. 2024. "The Future of Copyright and AI in the European Union: A Polish Perspective." European Journal of Law and Technology 15 (1): 123-140.

[87] Draft act amending the act on copyright and related rights and certain other acts, Council of Ministers (15 Feb 2024)

Polish AI developers may be at a competitive disadvantage compared to those in other EU member states if they have limited access to data for training AI models[88]. This may cause the number of Polish AI startups to drop and impede the expansion of the national AI market as a whole[89]. Furthermore, in order to fund AI development, news publishers throughout the EU are looking into ways to monetize TDM access. Polish publishers who completely withdraw from the market run the risk of losing out on possible revenue-sharing or licensing deals with AI developers who require access to copyrighted news content for training[90].

Outcomes

Poland's propose to exclude the creation of AI models from both Article 3 (research) and Article 4 (exceptions) of the EU Copyright Directive seems to be a misguided attempt that could impede the country's generative AI model development and use. The field of AI development research and innovation may suffer greatly as a result of this exclusion[91].

A significant outcome is the restriction of training data access. Large datasets are essential for training models in AI development. Article 3 restricts access to copyrighted materials for TDM, which may limit the kind and caliber of data Polish researchers can use to train their models[92]. In comparison to researchers in other EU states who have access to more extensive datasets, this restriction may make it more difficult for them to develop AI technologies that are competitive. This means that meeting with the quality of the data sets and as a result, the outcome of the model itself will be far more challenging and will put AI developers in a losing position.

Furthermore, this exclusion might hinder global cooperation in AI research[93]. International cooperation is common in AI research. Polish researchers may find it difficult to collaborate

[88] The Impact of Copyright Law on Artificial Intelligence Research (2023) by Daniel Gervais, Journal of Artificial Intelligence Law, Vol. 7, No. 1, pp. 1-42

[89] The State of Large Language Models 2022 (2022) by Patrick Heavener, Bard College

[90] The Text and Data Mining Exception and the Monetization of Text and Data (2021) by Eleftheria Marina Moschidou,European Intellectual Property Review, Vol. 53, No. 3, pp. 161-173

[91] The Impact of Copyright Law on Artificial Intelligence Research (2023) by Daniel Gervais, Journal of Artificial Intelligence Law, Vol. 7, No. 1, pp. 1-42

[92] The State of Large Language Models (2022) by Patrick Heavener, Bard College

[93] Global cooperation in artificial intelligence research (2020) by UNESCO Science Policy and Capacity Building Division, Science Policy and Governance

and share knowledge if they are unable to access copyrighted materials for TDM while their counterparts in other EU countries can. This will ultimately slow down the rate of innovation.

4.2 Bulgarian country study - A Delayed but Standard Implementation

State of implementaton of the new provisions

To implement the EU Copyright Directive (Directive (EU) 2019/790), Bulgaria took a more traditional approach, which stands in stark contrast to Poland's proposed exclusion. Bulgaria was one of the last member states to finally enact the Directive into national law, doing so more than two years after the implementation deadline, according to a December 2023 Kluwer IP Law blog post[94]. Bulgaria's implementation of TDM and news content is in line with the fundamental principles of the Directive, even with this delay[95].

Article 4(3) of the Directive's opt-out mechanism is kept in place as part of Bulgaria's strategy. By using machine-readable tags or other methods to indicate their opt-out status, news publishers can decide whether or not their content can be used for TDM activities. Furthermore, Bulgaria recognizes the TDM exceptions listed in Articles 3 and 4, which concern research organizations and institutions dedicated to cultural heritage, much like other EU members do. These exceptions strike a balance between the need to preserve intellectual property and the advantages of academic and cultural advancement by making copyrighted news content easier to access for legitimate research and preservation purposes.

Reaction

Though Bulgaria appears to be following the Directive in general, more investigation is required to ascertain the premises of its implementation[96]. This entails being aware of the particular protocols set up to allow publishers to decline participation in TDM activities, including any explicit policies and standardized means of informing opt-out status.[97].

[94] The last in line: Bulgaria implements the CDSM Directive (2023) by Kluwer Copyright Blog

[95] Implementation of the Directive on Copyright in the Digital Single Market (EU) 2019/790 in the Member States of the European Union(2022) by Edo IPR

[96] World Intellectual Property Organization (WIPO), Law on Copyright and Neighbouring Rights (Bulgaria)

[97] Finding Balance: Copyright and Text and Data Mining in the Digital Age (2019) by World Intellectual Property Organization (WIPO)

Outcomes

All things considered, Bulgaria's adoption of the EU Copyright Directive is a traditional but crucial step toward harmonizing domestic law with European standards and guaranteeing that the interests of researchers and content creators are fairly balanced within the legal framework.

4.3 Comparison and Possible solutions regarding the implementation of the provisions

4.4.1 Comparison of implementation of the Directive progress in Bulgaria and Poland

(1) How implementation of Article 4(3) influenced News Publishers in the country?

Bulgaria has fully implemented EU Copyright Directive and Follows the Directive's opt-out mechanism. News publishers can choose to opt-out of TDM through machine-readable tags. This allows them more control over content use but doesn't completely block access.

Meanwhile incorporation in Poland is still under debate, as it is not yet fully implenmented. Polish government excludes AI model creation from both Article 3 (research) and Article 4 exceptions. This restricts access to copyrighted materials for TDM, potentially hindering publishers' ability to monetize access through licensing deals.

(2) If the new Directive hasn’t implemented yet, how the country are planing to regulate TDM activities?

As the Polish implementation of EU Directive still under debate, the specific regulations and limitations regarding TDM activities in Poland remain unclear.

Copyright protection is currently based mostly on national law - Polish Copyright Act (Ustawa o prawie autorskim i prawach pokrewnych), defines the scope of copyright, ownership rights, limitations and exceptions, and enforcement mechanisms. Similar to the EU

Directive, copyright protection arises automatically upon creation of the work, without any need for registration.

(3) In both cases: How have news publishers reacted to the current state of copyright regulation for news articles?

Both publisher and AI-companies in Bulgaria are not yet sure how will the opt-out mechanism work in practice due to the luck of legal certainty and absence of technical protocols.

Meanwhile, Poland approach will make the country the market leader. It is also stated that Poland are working on #PLLUM project, creating a Polish language model, so if these regulations are passed, the Polish PLLUM model will speak in archaisms, after training in the public domain.

In summary, Bulgaria and Poland have contrasting approaches to text and data mining (TDM) within the EU Copyright Directive. Bulgaria's use of an opt-out mechanism empowers publishers while still allowing for research. On the other hand, Poland's exclusion of AI creation from TDM restricts innovation and research competition. Both countries face uncertainties: Bulgaria with unclear opt-out protocols and Poland with unproven benefits of their approach. These differing cases provide valuable insights for the EU on how to balance copyright protection and promote innovation in the digital age.

4.4.2 Possible solution to the challenges of implementation

There are alternative solutions Poland could consider to address its concerns while still fostering innovation: Nuanced Approach, Focus on Transparency and Traceability, Collaboration with the EU[98].

[98] Finding Balance: Copyright and Text and Data Mining in the Digital Age (2019) by World Intellectual Property Organization (WIPO)

Poland could also look into other options to address its issues and promote innovation at the same time. A more nuanced policy that permits TDM for research purposes under Article 3 with particular protections for commercially oriented AI development could be one strategy. When working with copyrighted materials, this may entail requiring researchers to obtain licenses or adhere to more stringent data usage limitations[99].

Furthermore, by emphasizing traceability and transparency, worries about how AI development may affect copyright may be allayed[100]. In order to ensure proper attribution to rightsholders, Poland may enact regulations requiring AI developers to disclose the source and usage of training data[101].

Interim conclusion

The application of EU Copyright Directive Article 4(3) in various Member States reveals a convoluted legal landscape with difficulties in real-world enforcement. Such a fragmented legal landscape is the result of the disparities in legal interpretation, which cause inconsistent application and enforcement of the opt-out provisions. The various methods used to incorporate the opt-out and TDM exception clauses show how coordinated strategies are required to preserve legal clarity and guarantee the efficient protection of copyright holders' rights.

Case studies from Bulgaria and Poland highlight the various ramifications and strategies for putting these provisions into practice, as well as the possible advantages and disadvantages of various strategies. The need for flexible legal frameworks and strong enforcement mechanisms is growing as the digital environment changes more and more. It will be essential to address these issues through cooperation, standardization, and creative solutions in order to strike a balance between news publishers' rights and the requirements of technological advancement and research.

[99] DiAngelo D. "Publishers Need an Opt-Out Strategy in 2023", Global Marketplace Development, Emodo, (Feb 2023)

[100] Responsible AI Development: A Framework for Stewardship (2020) by The Conference Board & McKinsey & Company

[101] Urs G., Silke E., "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age", working paper #2007-01 Berkman Center Research Publication

In summary, a complex interaction between legal requirements, technological capabilities, and news publishers' strategic interests is reflected in the practical application of Article 4(3). The application and upholding of these opt-out clauses must change in tandem with the ongoing evolution of the digital environment[102]. This continuous change emphasizes the need for flexible legal frameworks that can adjust to the changing needs of data use and copyright[103]. Legislators, publishers, and TDM practitioners will need to have ongoing discussions in order to develop such frameworks and guarantee that the rights of content creators and the advantages of technological advancements are protected and balanced.

[102] Copyright in a Digital Age (Third Edition) (2022) by Liisa Hyvönen and Christopher Millard, Oxford University Press

[103] The Future of Copyright: Balancing Creativity and Digital Innovation (2020) by Pamela Samuelson, Oxford University Press

V. Legal and Practical Challenges posed by provisions

This chapter explores the legal, technical, and operational challenges associated with the implementation of opt-out provisions under Article 4(3) of the EU Copyright Directive, emphasizing the significance of these provisions for news publishers. It also looks at the broader implications of these provisions on market dynamics, international cooperation, and the balance between copyright protection and access to information. By examining the theoretical and practical aspects of these opt-out provisions, we aim to provide a comprehensive understanding of their necessity and impact on the digital ecosystem.

4.1 Legislative challenges of implementation of European regulation

Balancing Copyright with Access to Information

The debate surrounding copyright law theory revolves around striking a balance between safeguarding the economic rights of copyright owners and promoting the unhindered flow of information. This debate is particularly relevant in the context of TDM[104]. Copyright regimes aim to protect the creative works of their creators, encouraging the production of original content. However, they must also take into account the public's interest in accessing and utilizing these works in new and innovative ways, such as through TDM for research, journalism, or technological development[105].

The balance between protecting copyright and enabling the dissemination of information is becoming more challenging in the digital age. With the ease and speed of copying and sharing information, the impact of copyright law can be significantly amplified through Text and Data Mining (TDM)[106]. This technology allows for the processing of vast amounts of data at unprecedented speeds and scales. The law must adapt to these technological capabilities without infringing on the fundamental rights of creators. The CDSMD Directive is an example

[104] Clark A., Calow D., "Training AI models: Consent., copyright and the EU and UK TDM exceptions", DLA Piper Blog

[105] Kristen G., et al "Training AI models on synthetic data: no silver bullet for infringement risk in the context of training AI systems (Part 3 of 4)", Claeary IP and Technology Insights

[106] Kaufman R., "Protecting Commercial AI Rights is harder than you think – EU Edition", Scholarly Kitchen

of a legal framework that aims to strike a balance by providing provisions for reasonable use while also allowing copyright holders to opt-out to manage the economic implications[107].

Upon analyzing these theoretical frameworks, it becomes evident that there is a complex relationship between innovation and protection[108]. Every modification in the law can have a significant impact on the digital media, research, and information dissemination environment. The ongoing challenge for lawmakers is to keep adapting these frameworks to keep up with technological advancements while ensuring that they remain fair and equitable for all stakeholders involved.

Lack of legal certainty of TDM regarding interpretation

The TDM exception for LLM creates some problems. The current structure is mainly intended for research activities conducted by non-profit institutions. There is ambiguity regarding the application of the TDM exemptions to commercial TDM activities carried out by individuals. In addition, LLMs must comply with the Terms of Use of the online content they access for TDM purposes[109].

The TDM exemption may not fully meet the needs of the LLM. According to the Directive, content used for TDM can only be stored for as long as necessary for the purposes of TDM. This means that AI developers may have to remove copyrighted content after the training phase, potentially excluding it from the validation or testing phases. There are also limits on the modification of content required to train artificial intelligence, and a three-step test ensures that copyright exceptions do not harm the interests of rights holders. These limitations highlight the need for a broader interpretation of TDM exceptions to cover the entire LLM development process, including the training, validation, and testing phases. Careful legal navigation is critical to supporting the sustainable development of AI technologies within copyright law[110].

[107] Paul Keller, "Generative AI and copyright: Convergence of opt-outs?" (Kluwer Copyright Blog, 2023)

[108] Laura L. "In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?" (Spawning Blog, 2024)

[109] Paul Keller, "Generative AI and copyright: Convergence of opt-outs?" (Kluwer Copyright Blog, 2023)

[110] Bernt Hugenholts, "The New Copyright Directive: Text and Data Mining (Articles 3 and 4)",)Kluwer Copyright Blog, 2019)

Potential need to broader Interpretation of TDM in a more normative manner[111]. This more expansive interpretation might cover the training, validation, and testing stages of AI development as well as the use of copyrighted content in all cases. To guarantee a sustainable course for AI development that upholds established copyright protections, careful legal navigation is necessary.

Each member state has its own definition of what constitutes sufficient notice for opting out[112]. Certain jurisdictions require highly specific and explicit notices to be attached directly to the content, while other jurisdictions allow notices to be more broadly applicable, covering entire websites or repositories. A fragmented legal landscape is the result of the disparities in legal interpretation, which also cause inconsistent application and enforcement of the opt-out provisions.

For policymakers, the ongoing discussion about generative AI models and the TDM exception poses a difficult task. Encouraging AI innovation while safeguarding copyright holders' rights will require careful consideration if the EU is to continue developing AI. Maintaining this equilibrium is crucial to ensuring that the digital era upholds the rightful rights of content creators while fostering innovation and the general good.

Ignoring the three-step test

It is not clear enough both will the TDM performed by AI-companies like ChatGPT, Gemini be considered as research and educational purpuses and will the answer differ if such actions were for commercial or non-commercial aims.

This blanket interpretation of the TDM exception should be reasonably considered as a failure to consider the three-step test[113] (and incompatible with the international obligations of the EU

[111] Rochelle C. Dreyfuss, Copyright and Innovation: The Struggle for Balance copyright and innovation the struggle for balance ON Oxford University Press (2017)

[112] The Text and Data Mining Exception and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732

[113] Paul Goldstein “Limitations and Exceptions to Copyright and Related Rights in the Digital Environment” Journal of Intellectual Property Law & Practice (2019) [Accessed 19 May 2024]

and of the Member States), which implies that no fair balance between the protection of copyright and related rights, on the one hand, and third-party rights and legitimate interests, on the other, may be fully achieved here. This has also led many to the conclusion that the Al Act should put an end to the discussion on the applicability of the TDM exception for purposes of generative Al[114].

However, EU courts are still bound to consider the three-step test in order to determine if the exception is available in the specific circumstances at hand, meaning that the acts of any defendant must satisfy the requirements of the three-step test. As such, the TDM exception cannot be regarded as having blanket applicability irrespective of whether the use in question is, for example, non-commercial or commercial.

The three-step test includes following steps:

Step 1	To determine the transformative nature of a use, the purpose and nature of the use are assessed, asking whether the use transforms the copyrighted material into something fresh and innovative.
Step 2	The type of copyrighted work—creative or factual—is considered, which is important in determining fair use.
Step 3	The impact of the use on the market is evaluated, taking into account the possibility that the use may harm the market of the original work.

The AI Act runs the risk of putting AI developers in an unbalanced position where they have more freedom to use copyrighted content without giving it enough thought in terms of how it might affect the rightsholder's market position or how transformative the use might be if it ignores this three-step test[115].

At the same time, copyright holders might be less motivated to produce new content if they are unable to effectively regulate the way in which their material is used for AI training[116]. Long-

[114] Copyright and the Digital Single Market Directive (2019) by Rochelle C. Dreyfuss, Oxford University Press

[115] Laura L. “In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?”(Spawning Blog, 2024)

[116] Patry on Copyright (Second Edition) (2019) by William Patry and Pamela Samuelson, Wolters Kluwer

term, this potential lack of control may impede innovation and creativity by reducing the financial incentives for producing original works.

By ignoring the three-step test, the AI Act risks putting AI developers in an unbalanced position where they have more freedom to use copyrighted content without paying enough attention to how it might affect the rights holder's market position or how transformative the use might be. if it ignores this three-step test.

Absence of a single standard mechanism for exercising the right of refusal

Looking critically, currently, there are only laws in the EU that provide for the author's right to refuse the use of his work in AI training, but there is no single mechanism or single technical protocol standard as such[117].

In practice, companies have found a way to ensure the right of refusal by expressing it in a form on the platform itself (ei Chat GPT).

Simplifying compliance and reducing the administrative burden for all parties involved can be achieved by implementing uniform and transparent opt-out procedures across the E[118]. This standardization would reduce legal ambiguities, provide clarity and consistency, and ensure that TDM practitioners and publishers are equally aware of the procedures and requirements. This strategy can improve the efficiency of cross-border TDM activities and ensure more seamless international cooperation.

Uncertainty regarding Data Base protection

The originality threshold is one prominent restriction[119]. Only databases that satisfy a particular standard of uniqueness in the choice and organization of their contents are covered by the Directive. Commonplace databases might not be eligible for this Directive's protection.

[117] Keller P., Warso Z. “Defining best practices for opting out of ML training”, Open Policy Brief (Sep 2023)

[118] Towards a Standardized Opt-Out Mechanism for Text and Data Mining in the EU by Max Planck Institute for Innovation and Competition

[119] EU Database Directive: A Commentary (2010) by Paul de Hert and Lilian Walker, Hart Publishing

Furthermore, the Database Directive follows the concept vs. expression distinction, just like copyright law does[120]. Not the underlying concepts or information in the news articles themselves, but the expression of the database's arrangement and selection is protected. Understanding the extent and constraints of the Directive's protection requires an awareness of this distinction.

Additionally, there is a relationship that occasionally overlaps between copyright law and the Database Directive[121]. Determining which right provides the best protection in a given circumstance is vital. For example, the Database Directive may provide better protection for the compilation of news articles than copyright for individual articles.

Need for international agreements

It can be challenging to enforce opt-out clauses, particularly when content is accessed from countries with different legal norms or by organizations without a direct legal obligation to abide by the EU directive[122]. These obstacles to enforcement highlight the need for stronger international agreements and collaboration in order to successfully protect these rights internationally. The protection of rights holders' interests is made more difficult by the lack of a cogent international enforcement mechanism, which also creates major obstacles to consistent enforcement.

5.2 Practical Challenges posed by the provisions

The opt-out provisions raise legal and ethical challenges regarding the fair use of copyrighted material for societal benefit, including academic and scientific research. There is ongoing

[120] Copyright Law (Third Edition) (2018) by Lionel Bently and Brad Sherman, Oxford University Press

[121] Copyright and Database Rights: A Comparative Analysis (2014) by Marie-Angèle Bouraoui, Edward Elgar Publishing

[122] The Future of International Copyright Enforcement (2022) by The Center for Intellectual Property & Information Technology Law, Fordham University School of Law ([cipit.law.fordham.edu]) (Accessed 19 May 2024)

debate over where the line should be drawn between protecting the economic rights of publishers and supporting public interest of research and development[123].

Implementing and enforcing opt-out provisions involves significant technical and operational challenges. Identifying and respecting opt-out notices requires advanced technology and coordination, which can be costly and complex, especially for smaller tech companies or academic institutions with limited resources[124].

Given the global nature of the internet and digital media, opt-out provisions in the EU may not be recognized or enforceable in other jurisdictions, leading to conflicts and legal uncertainties. This disparity can create a fragmented landscape where TDM practices must vary significantly from one country to another, complicating global operations for international news organizations and technology firms.

The opt-out provisions can also alter market dynamics by giving larger publishers more control over their content, potentially leading to market imbalances where only a few large entities have the power to set terms and conditions for TDM use. This could disadvantage smaller publishers or new entrants who might rely on more open access to digital content to compete.

While copyright encourages content creation, it also brings challenges in the digital age, especially regarding access and use of creative works for new purposes like TDM. Here's an overview of these key challenges:

No united technical protocol to perform right to opt-out

Currently EU only proposed the provisions, but no further guidance or technical solutions, protocols habe not been yet presented. That lead to making the provision incometable and even not possible to perform[125].

[123] Challenges and Opportunities for Text and Data Mining in the Scholarly Ecosystem by The Association of American Universities (AAU)

[124] The Future of Fair Use in the Digital Age by The Center for Democracy & Technology

[125] Keller P., “Generative AI and copyright: Convergence of opt-outs?”, Kluwer Copyright Blog (Nov 2023)

In practice, companies have found a way to ensure the right of refusal by expressing it in a form on the platform itself (eg Chat GPT). Other, like CoPilot by GitHub incorporated tool which shows from where the selected piece of information was taken.

Even though some steps have been taken by AI companies, such market distortion lead to disadvantaged positions for publishers, as they need to make a great effort to apply to each and every existing form to opt-out[126]. Such actions could be costy and at the same time impossible for small news organizations or individual authors due to the lack of financing and human resources[127].

Economic challenges

News publishers need control over their content in the digital age, where content is easily accessed and reused. Copyright law should empower them to decide if and how their content can be used for TDM activities and in that way allow publishers to monetize access to their content for TDM or negotiate licensing agreements reflecting the value of their works in data-driven markets.

News content creation requires significant investments in human resources, finances, and time. The current opt-out provision in some copyright frameworks helps protect these investments by preventing unauthorized use that could undermine potential revenue streams from their content.

Luck of interpretation

The TDM exception for LLMs poses specific challenges. The current framework is mainly intended for research activities conducted by non-profit institutions[128]. There is ambiguity regarding the application of TDM exceptions for commercial TDM activities carried out by

[126] Lopez-Tarruella A., "Google and Law. Empirical approaches to legal aspects of knowledge-economy business models", TMC Asser Press (2012), p. 113-168

[127] Kristen G., et al "Training AI models on synthetic data: no silver bullet for infringement risk in the context of training AI systems (Part 3 of 4)", Claeary IP and Technology Insights (Jan 2024)

[128] A Copyright Exception for Text and Data Mining (2020) LIBER Europe] [Accessed 19 May 2024]

private entities[129]. Additionally, LLMs must adhere to the Terms and Conditions associated with online content they access for TDM purposes.

The TDM exception may not fully meet the needs of LLMs. According to the Directive, content used for TDM can only be kept for as long as necessary for TDM purposes[130]. This means that LLMs might have to delete copyrighted content after the training phase, potentially excluding it from validation or testing phases. There are also restrictions on content modifications needed for AI training, and a three-step test ensures that copyright exceptions do not harm the rightsholders' interests These limitations highlight the need for a broader interpretation of TDM exceptions to cover the entire LLM development process[131], including training, validation, and testing phases. Careful legal navigation is crucial to support the sustainable development of AI technologies within the bounds of copyright law.

Interium conclusion

The relationship between copyright protection and access to information for Text and Data Mining (TDM) activities is a complex challenge in the digital age. The EU's Copyright Directive in the Digital Single Market (DSM) aims to strike a balance by providing a TDM exception with limitations and opt-out mechanisms. However, there are still significant challenges in implementing and interpreting this directive.

As for the legal challenges the most significant are: Luck of Interpretation of TDM Exception for LLMs, and no international cooperation to agree upon one standard protocol for performing opt-out.

[129] Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue? (2020) by Bernt Hugenholtz SSRN [Accessed 19 May 2024]

[130] Directive 2001/61/EC of the European Parliament and of the Council of 8 May 2001 on Copyright in the Information Society [Accessed 19 May 2024]

[131] Urs G., Silke E., "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age", working paper #2007-01 Berkman Center Research Publication

The TDM exception's applicability to commercial LLM activities should be clarified to address any ambiguity. Moreover, restrictions on content storage and modification could impede the entire LLM development process. A more comprehensive interpretation that includes training, validation, and testing phases could be advantageous
Additionally, opt-out clauses are not highly effective when content is accessed from countries with different legal frameworks. Therefore, stronger international cooperation and harmonization are crucial for robust enforcement.

As for the practical challenges the most would namely be impossibility no protect copyright due to the absanse of united protocol and economical challenges both for news publishers and AI-companies.

As news publishers and authors receive their income by having control of the data available and simplified opt-out mechanisms should be proposed. As for AI-companies so monetieze their product, they need to create a smart AI-model, that is capable of performing variety of difficult tasks and huge number of quality data is need to teach model to do so.

VI. Recommendation of Compliance strategies

Navigating the complexity of EU Copyright Directive 2019/790 (Article 4(3)) can be a difficult task for new organizations entering the digital landscape. This chapter offers new organizations a thorough manual for creating compliance strategies that work and for comprehending the significance of protocol alignment for the success of Text and Data Mining (TDM) operations.

6.1 Possible steps to be considered by EU goverment

Unite implementation

The problem of legal disparity might potentially be resolved by attempts to standardize copyright regulations and enforcement procedures among various authority[132]. The fragmentation that currently results from different national implementations of Article 4(3) of the EU Copyright Directive would be lessened by global harmonization, which would guarantee more level playing fields for TDM practices. This strategy could improve cross-border TDM activities' efficacy and enable more seamless international collaborations[133].

United Protocol and relevant Tools

Simplifying compliance and lowering administrative burdens for all parties involved could be achieved by implementing uniform and transparent opt-out procedures throughout the EU[134]. Legal ambiguities would be reduced by this standardization, which would ensure clarity and consistency and make sure that TDM practitioners and publishers alike are aware of the procedures and requirements.

[132] Towards a Standardized Opt-Out Mechanism for Text and Data Mining in the EU by Max Planck Institute for Innovation and Competition

[133] European Parliament, Directorate General for Internal Policies (2018) 'The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects'

[134] The Need for Standardized Opt-Out Mechanisms in Text and Data Mining by European Digital Rights (EDRi)

There should be created the Tools for Data compliance in order to guarantee that opt-out notices are honored during the data mining process, these tools are essential to ease of use and efficacy of the opt-out infrustructute[135].

Propose licensing schems

Another workable option would be to investigate alternative licensing schemes that reward responsible TDM practices and pay rightsholders for the value of their data[136]. These models might provide a reasonable solution that would both ensure that news publishers receive just compensation and permit TDM activities to flourish[137]. In order to promote a win-win partnership, licensing agreements could be customized to represent the unique requirements and contributions of both parties[138].

Avoiding monopolization

Companies that collect and process large amounts of data may have a competitive advantage over their competitors who do not have access to such data. At the same time, the development and use of AI algorithms may lead to anti-competitive behavior, such as price discrimination or predatory pricing.

In the US[139], there are two main antitrust laws that can be applied to cases where the use of AI leads to anti-competitive behavior, such as data monopolization. The Sherman Act prohibits monopolization and conspiracy in restraint of trade (Section 1) and prohibits abuse of a dominant market position (Section 2).

Encouraging collaboration between authors and AI developers

The South Korean government is taking a number of measures, recognizing the importance of pooling their knowledge and experience to develop a responsible and innovative ecosystem, including the following:

---

[135] Machine-Readable Copyright Information: A Study on Legal Frameworks and Technical Standards (2023) by World Intellectual Property Organization (WIPO)

[136] The Need for Standardized Opt-Out Mechanisms in Text and Data Mining by European Digital Rights (EDRi)

[137] The Challenges of Global Copyright Harmonization in the Digital Age by University of California, Los Angeles (UCLA)

[138] The Impact of the EU Copyright Directive on Text and Data Mining (2020) by A. Liogier

[139] The Future of Fair Use in the Digital Age by The Center for Democracy & Technology

• Funding research aimed at studying the ethical and legal aspects of AI, as well as for the development of new tools and methods that contribute to the protection of copyright.
• Educational activities to promote better understanding and cooperation between the two groups, and to help them build connections and partnerships.
• Creation of platforms and infrastructure where AI authors and developers can communicate, share experiences and collaborate on projects.

Ensuring database security

In South Korea, the law on Special Information Industries, which regulates the protection of databases, has been updated, setting requirements for their collection, use and disclosure.

Provisions regarding the use of databases for machine learning focus on such aspects as: classification of databases, access and use control, privacy protection, implementation of technical security measures of databases against unauthorized access and use[140].

Encryption and anonymization

The South Korean government is also investing in research and development of new encryption methods to protect data during storage and transmission, as well as in the development of anonymization techniques[141] that will allow AI models to be trained without revealing personal information.

Using special TDM tools that involve minimally copying a few words or scanning data and processing each item separately is the middle ground. This approach will provide sufficient quality data for training and at the same time pose minimal risks of copyright infringement[142].

[140] Moschidou, E. M. (2021). The Text and Data Mining Exception and the Opt-Out Mechanism under the EU Copyright Directive. European Intellectual Property Review, 53(11), 721-732.

[141] Can Text and Data Mining Thrive in the Shadow of Opt-Out? by University of Amsterdam

[142] Machine-Readable Copyright Information: A Study on Legal Frameworks and Technical Standards (2023) by World Intellectual Property Organization (WIPO)

Consider allowing TDM for non-commercial purposes with any restrictions

The UK government will allow the use of TDM for research purposes as long as it is of a non-commercial nature. It is also noted that it is a prerequisite that the beneficiaries have legal access to the information[143].

Explaining its position, the government stated: "Copying related to text and data analysis is a necessary part of the technological process and is unlikely to replace relevant work (such as a journal article). Therefore, it is unlikely that permission to mine for research will in itself adversely affect the market or value of copyrighted works. Indeed, it is possible that removing the restrictions on analytical technologies will increase the value of articles to researchers."

6.2 Compliance Strategies for AI-companies

Waitig for a common technical protocol at the European or global level, AI developers can comply with the provisions of the Copyright Directive by creating a form on their platform to refuse or contest the use of the work for the purpose of training an AI model[144].

From previous global research, it became known that the use of special TDM tools, which involve minimal copying of a few words or scanning of data and processing of each element separately, minimizes the possibility of copyright infringement.

At the same time, it should be taken into account that such a practice does not put Ukrainian developers of AI models in a competitive disadvantage if other players on the market will use larger and better data sets for their training in order to circumvent copyright[145].

In addition, AI developers can create or review an already developed knowledge base on the development of methods for anonymous training of AI models.

---

[143] A Global Approach to Text and Data Mining by The National Academies Press https://www.science.org/doi/10.1126/science.add6124

[144] The Future of Copyright: Balancing Creativity and Digital Innovation (2020) by Pamela Samuelson, Oxford University Press

[145] Legal and Technical Interoperability Challenges for Text and Data Mining (2019) by S. Ruegge

6.3 Complience strategies for news publishers and authors

First and foremost, new organizations need to be fully aware of the legal context of Article 4(3) and how it affects TDM procedures. This involves comprehending the reach of opt-out clauses, which calls for a precise knowledge of the kinds of information and actions that fall under the purview of the opt-out process. This information aids in determining whether a particular piece of content needs to have its opt-out notices checked before TDM activities can continue[146].

Organizations also need to take into account the EU-only territorial scope of Article 4(3). In order to ensure wider compliance for TDM activities with a global reach, it is imperative to take copyright laws and opt-out mechanisms in other jurisdictions into account. Furthermore, it's critical to comprehend the fair use exceptions incorporated into EU national copyright laws. Even in the presence of an opt-out notice, fair use exceptions may offer alternate paths for legal TDM practices, depending on the nature of the TDM activity and the kind of content used[147].

In order to reduce the possibility of legal issues, new organizations must also take a proactive approach to risk management. It can be advantageous to conduct regular legal audits to evaluate compliance procedures and spot any gaps. These audits help reduce the legal risks associated with non-compliance and enable course correction[148]. Furthermore, investigating specialist insurance plans made to address possible copyright violations resulting from TDM operations might be a wise way to reduce risk.

Currently, copyright holders have only two options: to allow free use of their works or to refuse to grant permission for such use. The middle is not given. An alternative win-win solution would be to allow the introduction of some payment (premium packages for researchers or licensing) for the use of works to train AI models.

[146] The Impact of the EU Copyright Directive on Text and Data Mining (2020) by A. Liogier

[147] Challenges and Opportunities for Text and Data Mining in the Scholarly Ecosystem by The Association of American Universities (AAU) https://www.aau.edu/research/featured-research

[148] The Impact of Text and Data Mining on Innovation by Center for Information Policy Leadership (CIPL) at Hunton Andrews Kurth LLP

The first step for rights holders is to implement technical measures to secure databases against unauthorized access and use. Having ensured control over access to databases, there will be an opportunity for monetization - providing access to data or part of it for a financial reward.
Copyright law should allow publishers to monetize access to their content for TDM or enter into licensing agreements that reflect the value of their works in data-driven markets.

In this way, Publishers could create specialized products for different market segments and differentiate their offerings accordingly. For example, they could use their control of TDM to create premium content packages for researchers or companies, and then use that control to negotiate better terms.

Another viable option would be to explore alternative licensing schemes that reward responsible TDM practices and pay rights holders for the value of their data.

6.4 Progres in solving the Challenge of Diverse Technical Protocols

While some online materials are available under permissive licenses that allow for reuse, others are protected by website Terms and Conditions that prohibit web scraping, creating a complex legal landscape for LLM training. However, the implementation of these opt-outs, particularly in machine-readable formats, remains unclear. For instance, OpenAl's GPTBot can be blocked via a site's robots.txt. So, in Europe, TDM is effectively opt-out[149]. An important obstacle that both new and established organizations must overcome is the lack of standardized technical protocols for sending Article 4(3) opt-out notices.

This lack of standardization presents a number of challenges[150].
First off, smaller organizations with fewer resources may find it especially difficult to meet the technical requirements of various opt-out protocols due to their complexity[151].

[149] Legal and Technical Interoperability Challenges for Text and Data Mining (2019) by S. Ruegge

[150] Challenges in Implementing the Text and Data Mining Exception in the EU by U Putra & H Liu (2021)

[151] Text and Data Mining in the Context of Scholarly Publishing (2020) by Laura Quilter, Journal of Librarianship and Scholarly Communication, Vol. 8, No. 1

Second, TDM practitioners find it challenging to develop universal strategies for recognizing and honoring these notices on a variety of platforms due to the uneven application of opt-out mechanisms by different publishers.
Lastly, it is an expensive undertaking to create and maintain systems that can adjust to different opt-out protocols, particularly for new organizations with tight budgets[152].

As a solution currently there are two most significant proposed protocols "Wide Web Consortium" (W3C) by Text and Data Mining Reservation Rights Community Group and "Baseline report of policies and barriers of TDM in Europe" by FutureTDM.

The **World Wide Web Consortium (W3C)** formed the Text and Data Mining Reservation Rights Community Group in response to the need for standardization.

This group is essential to the development of a standardized technical framework that would make it easier for publishers and TDM practitioners to comply with opt-out notices. Standardized protocols promote interoperability, making it possible for TDM tools and systems to recognize and honor opt-out notices on any platform with ease. By offering a transparent and uniform framework, this initiative greatly lessens the compliance burden for new organizations joining the TDM field.

For news publishers, implementing a standardized framework would ensure that their opt-out notices are clear and consistently enforced across the web. This clarity would strengthen their control over their content and potentially increase the effectiveness of their opt-out strategy. Additionally, a transparent and uniform framework would reduce the burden for new organizations entering the TDM field. By simplifying compliance procedures, the W3C's initiative could encourage broader participation in TDM research and development, ultimately benefiting both news publishers and the TDM community.

The FutureTDM "Baseline report of policies and barriers of TDM in Europe" provides insightful information about the status of TDM practices in Europe today, emphasizing that a major obstacle to the wider adoption of TDM activities is the existence of disparate technical protocols.

---

[152] The Impact of the EU Copyright Directive on Text and Data Mining (2020) by A. Liogier

In order to promote a more effective TDM ecosystem in Europe, the report stresses the need for active support of the W3C's efforts to develop standardized protocols. Additionally, the report emphasizes how important it is to educate all parties involved—publishers, TDM practitioners, and legislators—about the advantages of standardization[153]. A consensus on standardized protocols can only be reached by encouraging collaboration between all parties, which can be accomplished through open dialogue, workshops, and other cooperative efforts.

Main difference is that the W3C prioritizes technical solutions (standardized framework) for declaring opt-out status, while The FutureTDM report emphasizes the broader compliance landscape and the importance of clear communication between stakeholders.

From the analyses made we can see the following:

Declaration of TDM Reservation Rights - While some materials have permissive licenses, others lack clear opt-out mechanisms, especially in machine-readable formats. This creates a complex legal landscape for both publishers and TDM practitioners. For example, some opt-out mechanisms rely on website Terms and Conditions prohibiting web scraping, which can be unclear and inconsistently applied.

Expressing a TDM Policy - A documented TDM policy by news organizations clarifies their approach to data access and permissions.This transparency fosters collaboration with TDM practitioners who can better understand specific opt-out procedures and potential licensing opportunities. Open communication can lead to "win-win" agreements and reduce potential disputes.

Stakeholder Policies - News organizations should consider the policies of rights collectives and individual authors when developing their TDM strategies. This fosters a more comprehensive view of the compliance landscape. Building relationships with relevant industry associations and legal advocacy groups can provide valuable information on legal updates, best practices, and networking opportunities.

[153] Hugenholtz, P Bernt (2020) 'Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?' SSRN

While the W3C and FutureTDM take slightly different approaches, they are complementary. The W3C's standardized framework can address the technical challenges of opt-out notices, while the FutureTDM report's emphasis on communication encourages a more holistic view of TDM compliance that considers all stakeholders.

Interium conclusion

Article 4(3) can be challenging for new organizations to navigate, but it is still possible to do so by putting proactive compliance strategies into place and keeping up to date with developments.

As government plays key role in creating regultions and recommendations it is essential that they consider to provide wider application of the provisions s well as working solution oninternational level.

New organizations can create strong compliance strategies by taking a multifaceted approach that includes technical implementation, legal knowledge, risk management techniques, and cooperative relationship building. This all-encompassing strategy puts newcomers in a successful position in the rapidly changing digital market while guaranteeing responsible TDM practices within the parameters of Article 4(3).

To mitigate legal risks, LLMs are advised to autonomously analyze website Terms and Conditions to distinguish between materials that are freely usable for training and those that are not. The OpenAl GPTbot web crawler, allowing website owners to opt-out or filter content access, is a significant step towards addressing IP law concerns, potentially setting a future standard for LLM providers.

VII.Conclusion

7.1 Research Outcomes

The complex nature of Article 4(3) of the EU Copyright Directive and its consequences for Text and Data Mining (TDM) practices—particularly with regard to news content—have been examined in this thesis.

The study produced a number of noteworthy results:

First, despite the conflict of copyright protection, technical innovation with fundamental right to access the information, news organixations fall within the scope of Article 4 (3) of the EU Copyright Directive.

However, there is still legal uncertainty surrounding the interpretation of rights under Article 4(3). EU government should provide abetter understanding how the opt-out notice applies to various content and activity types is essential for figuring out how it works in particular TDM scenarios. To guarantee that TDM practitioners can successfully navigate the legal landscape, this clarification is crucial.

Secondly, there are significant concerns regarding how Article 4(3)'s territorial reach will affect TDM operations involving content from non-EU jurisdictions. The provisions of the Directive may have far-reaching effects due to the global nature of digital content and TDM practices, which calls for a more thorough examination of international copyright laws and how they affect TDM operations.

It is clear that different member states have different implementation strategies for the opt-out clauses. While some nations, like Poland, suggest significant exclusions, others, like Bulgaria, have embraced more standardized methods. This discrepancy suggests that the EU's legal system is fragmented, which could make it more difficult to implement TDM practices.

Thirdly, the study shows that both AI-companies and news organizations face significant obstacles. Significant operational and technical challenges arise from the lack of standardized technical protocols for opt-out notices, particularly for smaller organizations. This

lack of consistency can make it more difficult for new players to enter the market and hinder the effective operation of TDM operations. Moreover, The ambiguity surrounding the applicability of the TDM exception to AI model training requires clarification.

As in practice, no official guidelines on technical solutions for enabling the right of reservation for creators, led to companies like Google, OpenAi, and Microsoft caming up with their opt-out tools and protocols to comply with the new regulations. As a result, a model-specific opt-out mechanism is worthless and costly, as creators are required to repeatedly provide opt-outs for each entity that trains models, which would consume disproportionate resources, especially for Digital Newspapers.

Last but not least, even though Article 4(3) seeks to achieve a balance between conflicting rights, there are still some risks associated with upholding these rights.

There's a chance that the opt-out process will disproportionately benefit big publishers. This might cause an imbalance in the market where these publishers set the terms for TDM access, stifling the competition from smaller players. A situation like this could discourage creativity and narrow the range of products available.

The legal environment surrounding TDM activities is fragmented due to the various national implementations of Article 4(3). Organizations that operate internationally face serious legal uncertainties as a result of this fragmentation. It can be difficult to navigate these differences, especially for smaller organizations that don't have the resources to handle intricate legal compliance in several jurisdictions.

Interpreting opt-out clauses too narrowly may discourage research and innovation. These interpretations run the risk of undermining research efforts by restricting access to important data for TDM activities, both commercial and academic. In order to promote a dynamic and creative research environment, it is imperative that the balance of rights does not unnecessarily restrict data access.

The study also highlights the continuous conflict between promoting innovation through TDM and defending the financial rights of news publishers. To protect their financial interests, publishers must have control over the content they publish, which is made possible by the opt-out clauses. But these rules also have the potential to hinder digital-age research and

development, which makes striking a balance between rights and innovation extremely difficult.

Thesis has contributed to the knowledge base in following points:

- It provides a comprehensive analysis of the legal and practical implications of Article 4(3) for TDM practices in the EU.
- The comparative country studies on Poland and Bulgaria offer valuable insights into the diverse approaches to implementing the TDM exception.
- It provides recommendations and compliance strategies for AI-companies and news organzations, as well as possible updates and variations to regulation for government consideration.

7.2 Final Thoughts

A vital discussion regarding striking a balance between news publishers' rights and the necessity of promoting text and data mining (TDM) for innovation and the public interest has been spurred by the implementation of Article 4(3). The complexity of the situation and the need for more work to create a more effective and balanced TDM ecosystem have been brought to light in this thesis. Numerous viable remedies have been recognized:

- Standardized Opt-Out Protocols: As recommended by the W3C Text and Data Mining Reservation Rights Community Group, the development of standardized technical protocols for opt-out notices would streamline compliance and improve effectiveness for all parties involved. TDM practitioners' current operational difficulties would be lessened by this standardization, especially for smaller businesses.
- Harmonization: In order to create a more level playing field and lessen legal uncertainty for businesses operating globally, efforts should be made both within and outside of the EU to harmonize copyright laws and enforcement procedures pertaining to TDM. A unified strategy would help to reduce the disarrayed legal environment and offer more precise direction for TDM operations.

- Alternative Models of Licensing: Investigating alternative models of licensing that reward responsible TDM practices and pay rightsholders for the value of their data may provide a workable way to support innovation while upholding publisher rights. These models would strike a balance between the need for data access for research and development and economic interests.

A more impartial interpretation and application of the TDM opt-out provisions can be accomplished by encouraging cooperation amongst legislators, rightsholders, tech firms, and researchers. This will guarantee that, while upholding the rightful rights of content creators, the digital age will continue to promote innovation and the general good.

The future of TDM and copyright law will likely be characterized by ongoing adjustments and refinements as both legal systems and technological capabilities evolve. The goal will be to find a balance that respects the rights of copyright holders while fostering an environment conducive to innovation and access to information.

Bibliography

Promary literature

1. 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 (CDSMD), art 4(3) (date of access: 28 February 2024).
2. Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (date of access: 28 February 2024).
3. 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, art 5.3c. (date of access: 28 February 2024).

Secondary literature

1. "Appeal for action on violations of the Berne Convention by the application to copying of creative works for AI development of the TDM exception in Articles 3 and 4 of the 2019 EU Directive on Copyright" by Authors Coalition (July 2023) <https://mse.dlapiper.com/post/102ivrx/training-ai-models-content-copyright-and-the-eu-and-uk-tdm-exceptions> accessed on 26 February 2024
2. "Baseline report of policies and barriers of TDM in Europe", FutureTDM, Horizon 2020 <https://pure.uva.nl/ws/files/8885048/FutureTDM_D3.3_Baseline_Report_of_Policies_and_Barriers_of_TDM_in_Europe.pdf> accessed on 3 February 2024
3. Bertuzzi L. "AI Act: EU Commission attempts to revive tiered approach shifting to General Purpose AI", EURACTIV (Nov 2023) https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-commission-attempts-to-revive-tiered-approach-shifting-to-general-purpose-ai/ accessed on 19 January 2024
4. Bernt Hugenholts P., "The New Copyright Directive: Text and Data Mining (Articles 3 and 4)", Kluwer Copyright Blog (July 2019

5. https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ accessed on 18 January 2024
6. Bently, Lionel & Sherman, Brad. (2018). Copyright Law (Third Edition). Oxford University Press.
7. Bouraoui, Marie-Angèle. (2014). Copyright and Database Rights: A Comparative Analysis. Edward Elgar Publishing.
8. Clark A., Calow D., "Training AI models: Consent., copyright and the EU and UK TDM exceptions", DLA Piper Blog (Jan 2023) https://mse.dlapiper.com/post/102ivrx/training-ai-models-content-copyright-and-the-eu-and-uk-tdm-exceptions accessed on 19 February 2024
9. The Center for Intellectual Property & Information Technology Law, Fordham University School of Law ([cipit.law.fordham.edu]) (Accessed 19 May 2024).
10. Conference Board & McKinsey & Company. (2020). Responsible AI Development: A Framework for Stewardship.
11. DiAngelo D. "Publishers Need an Opt-Out Strategy in 2023", Global Marketplace Development, Emodo, (Feb 2023) https://advertisingweek.com/publishers-need-an-opt-out-strategy-in-2023/ accessed on 18 February 2024
12. Dreyfuss, Rochelle C. (2017). Copyright and Innovation: The Struggle for Balance. Oxford University Press.
13. European Digital Rights (EDRi). The Need for Standardized Opt-Out Mechanisms in Text and Data Mining.
14. European Parliament, Directorate General for Internal Policies, Policy Department A: Citizens' Rights and Constitutional Affairs (2018). The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects. [Accessed 19 May 2024]
15. FutureTDM Baseline Report (year needed). Baseline report of policies and barriers of TDM in Europe.
16. Górski, Jan M. (2023). "The Copyright Directive and Text and Data Mining in Poland: Balancing Interests in the Digital Age." Journal of Intellectual Property Law & Practice 18(6): 489-502.
17. Hyvönen, Liisa & Millard, Christopher. (2022). Copyright in a Digital Age (Third Edition). Oxford University Press.
18. Kaufman R., "Protecting Commercial AI Rights is harder than you think – EU Edition", Scholarly Kitchen (Feb 2024)

https://scholarlykitchen.sspnet.org/2024/02/01/protecting-commercial-ai-rights-is-harder-than-you-think-eu-edition/ accessed on 11 February 2024

19. Keller P., "Generative AI and copyright: Convergence of opt-outs?", Kluwer Copyright Blog (Nov 2023) https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs accessed on 2 March 2024
20. Keller P., Warso Z. "Defining best practices for opting out of ML training", Open Policy Brief (Sep 2023) https://openfuture.eu/wp-content/uploads/2023/09/Best-_practices_for_optout_ML_training.pdf accessed on 14 February 2024
21. Keller P., "TDM: Poland challenges the rule of EU copyright law", Kluwer Copyright Blog (Feb 2024) https://copyrightblog.kluweriplaw.com/2024/02/20/tdm-poland-challenges-the-rule-of-eu-copyright-law/ accessed on 22 February 202
22. Kristen G., et al "Training AI models on synthetic data: no silver bullet for infringement risk in the context of training AI systems (Part 3 of 4)", Claeary IP and Technology Insights (Jan 2024) <https://www.clearyiptechinsights.com/2024/01/training-ai-models-on-synthetic-data-no-silver-bullet-for-ip-infringement-risk-in-the-context-of-training-ai-systems-part-3-of-4/#_ftn5> accessed on 20 January 2024
23. Lazarova A. "The last in line: Bulgaria implements the CDSM Directive", Kluwer Copyright Blog (Dec 2023) https://copyrightblog.kluweriplaw.com/2023/12/27/the-last-in-line-bulgaria-implements-the-cdsm-directive/ accessed on 25 February 2024
24. Laura L. "In the EU, Opt-outs Are the Way Forward? What the EU's TDM copyright exceptions mean for researchers, developers and rights holders?", Spawning Blog (Feb 2024) https://spawning.substack.com/p/in-the-eu-opt-outs-are-the-way-forward accessed on 28 February 2024
25. Lopez-Tarruella A., "Google and Law. Empirical approaches to legal aspects of knowledge-economy business models", TMC Asser Press (2012), p. 113-168, accessed on 15 February
26. Mezei P., "A saviour or a dead end? Reservation of rights in the age of generative AI", Social Science Research Network, p.1-13 (February 2023) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4695119 accessed on 22 March 2024
27. Max Planck Institute for Innovation and Competition. (year needed). Towards a Standardized Opt-Out Mechanism for Text and Data Mining in the EU.

28. Moschidou, Eleftheria Marina. (2021). "Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive." European Intellectual Property Review 53(11): 721-732.
29. Patry, William & Samuelson, Pamela. (2019). Patry on Copyright (Second Edition). Wolters Kluwer.
30. Samuelson, Pamela. (2020). The Future of Copyright: Balancing Creativity and Digital Innovation. Oxford University Press.
31. Twardowski, Piotr. (2024). "The Future of Copyright and AI in the European Union: A Polish Perspective." European Journal of Law and Technology 15(1): 123-140.
32. University of California, Los Angeles (UCLA). The Challenges of Global Copyright Harmonization in the Digital Age.
33. Urs G., Silke E., "EUCD Best Practice Guide: Implementing the EU Copyright Directive in the Digital Age", working paper #2007-01 Berkman Center Research Publication <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=952561> accessed on 14 February 2024
34. World Wide Web Consortium (**W3C**) 'Text and Data Mining Reservation Rights Community Group' (Feb 2022) < https://www.w3.org/2022/tdmrep/> accessed on 23 February 2023
35. World Intellectual Property Organization (WIPO). (2023). "What is copyright?" [Accessed June 18, 2024]
36. World Intellectual Property Organization (WIPO). (2023). Machine-Readable Copyright Information: A Study on Legal Frameworks and Technical Standards.
37. Hugenholtz, P Bernt (2020) 'Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?' SSRN [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4134651 [Accessed 19 May 2024]
38. European Parliament, Directorate General for Internal Policies (2018) 'The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects' [online] Available at: https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA%282018%29604941_EN.pdf [Accessed 19 May 2024]
39. LIBER Europe (2020) 'A Copyright Exception for Text and Data Mining' [online] Available at: https://libereurope.eu/document/a-copyright-exception-for-text-and-data-mining-2/ [Accessed 19 May 2024]

40. University of Turku (2023) 'Text and data mining (TDM) in the EU: What you need to know about copyright law and data analysis' [online] Available at: https://utuguides.fi/researchdata/instructions [Accessed 19 May 2024] (
41. The Copyright Alliance (n.d.) 'Copyright and Journalists' [online] Available at: https://copyrightalliance.org/ [Accessed 19 May 2024]
42. Directive 2001/61/EC of the European Parliament and of the Council of 8 May 2001 on Copyright in the Information Society (https://eur-lex.europa.eu/eli/dir/2019/790/oj)
43. Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue? (2020) by P Bernt Hugenholtz SSRN [Accessed 19 May 2024]
44. Directive 2001/61/EC of the European Parliament and of the Council of 8 May 2001 on Copyright in the Information Society (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32001L0061)
45. Case C-433/09 – Newspaper Licensing Agency Ltd v Meltwater Holding BV [2011] ECR I-012131 (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A62013CJ0360)
46. The Copyright Alliance, "Copyright and Journalists", https://copyrightalliance.org/ [Accessed 19 May 2024]
47. EU Database Directive: A Commentary (2010) by Paul de Hert and Lilian Walker, Hart Publishing
48. Copyright Law (Third Edition) (2018) by Lionel Bently and Brad Sherman, Oxford University Press
49. Copyright and Database Rights: A Comparative Analysis (2014) by Marie-Angèle Bouraoui, Edward Elgar Publishing
50. The Digital News Report 2023 (2023) by Reuters Institute for the Study of Journalism (University of Oxford)
51. Paul J. Heald, opyright and the Financing of Journalism ([invalid URL copyright and the financing of journalism, Oxford University Press (2016)
52. Poynter Institute, Journalism Ethics [Accessed 19 May 2024]
53. Lih-Fen Lin et al., Copyright in a Digital Age, Oxford University Press (2018)
54. Yannis Manolopoulos et al., Text and Data Mining for the Social Sciences: Big Data Applications in Research, Edward Elgar Publishing (2017)
55. Arvind Narayanan, Bias in Big Data, Springer (2017)

56. Rochelle C. Dreyfuss, Copyright and Innovation: The Struggle for Balance ([invalid URL copyright and innovation the struggle for balance ON Oxford University Press (2017)
57. D. W. Milligan, Newspapers in a Digital Age ([invalid URL newspapers in a digital age, Cambridge University Press (2017)
58. [1] Jessica Litman, The Text and Data Mining Exception in US Copyright Law, (2013) by Jessica Litman, Stanford Law Review, Vol. 66, No. 2, pp. 361-465 (https://crln.acrl.org/index.php/crlnews/article/view/24383/32222) (2013)
59. The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (2018) by Paul Keller et al., European Parliament, Directorate General for Internal Policies, Policy Department A: Citizens' Rights and Constitutional Affairs
60. Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732
61. Copyright and the Digital Single Market Directive (2019) by Rochelle C. Dreyfuss, Oxford University Press
62. Patry on Copyright (Second Edition) (2019) by William Patry and Pamela Samuelson, Wolters Kluwer
63. Rochelle C. Dreyfuss, Copyright and Innovation: The Struggle for Balance copyright and innovation the struggle for balance ON Oxford University Press (2017)
64. EU AI Act: shaping Copyright compliance in the age of AI Innovation (2024) by KEA European Affairs
65. The Future of Intellectual Property Law in the Artificial Intelligence Era (2020) by Florian Strenger, Edward Elgar Publishing
66. The Database Directive and News Aggregation: Striking a Balance Between Rights and Innovation (2013) by Martin Kilian et al., European Journal of Law and Information Technology, Vol. 6, No. 1, pp. 1-22
67. Text and Data Mining Rights and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732
68. The Challenges of Copyright Law Harmonization in the Digital Single Market (2018) by Eleftheria Marina Moschidou,European Intellectual Property Review, Vol. 50, No. 12, pp. 837-847

69. The Text and Data Mining Exception in the EU Copyright Directive: Best Practices for Implementation (2020) by COMMUNIA Association
70. The Impact of Copyright Law on Artificial Intelligence Research(2023) by Daniel Gervais, Journal of Artificial Intelligence Law, Vol. 7, No. 1, pp. 1-42
71. The State of Large Language Models 2022 (2022) by Patrick Heavener, Bard College
72. Global cooperation in artificial intelligence research (2020) by UNESCO Science Policy and Capacity Building Division, Science Policy and Governance
73. The Impact of Copyright Law on Artificial Intelligence Research (2023) by Daniel Gervais, Journal of Artificial Intelligence Law, Vol. 7, No. 1, pp. 1-42
74. The State of Large Language Models 2022 (2022) by Patrick Heavener, Bard College
75. The Text and Data Mining Exception and the Monetization of Text and Data (2021) by Eleftheria Marina Moschidou,European Intellectual Property Review, Vol. 53, No. 3, pp. 161-173
76. Finding Balance: Copyright and Text and Data Mining in the Digital Age (2019) by World Intellectual Property Organization (WIPO)
77. Responsible AI Development: A Framework for Stewardship (2020) by The Conference Board & McKinsey & Company
78. The last in line: Bulgaria implements the CDSM Directive (2023) by Kluwer Copyright Blog ([copyrightblog.kluweriplaw.com]) (Accessed 19 May 2024)
79. Implementation of the Directive on Copyright in the Digital Single Market (EU) 2019/790 in the Member States of the European Union(2022) by Edo IPR
80. World Intellectual Property Organization (WIPO), Law on Copyright and Neighbouring Rights (Bulgaria) ([wipolex-res.wipo.int]) (Accessed 19 May 2024)
81. Finding Balance: Copyright and Text and Data Mining in the Digital Age (2019) by World Intellectual Property Organization (WIPO)
82. The Future of International Copyright Enforcement (2022) by The Center for Intellectual Property & Information Technology Law, Fordham University School of Law ([cipit.law.fordham.edu]) (Accessed 19 May 2024)
83. The Impact of Text and Data Mining on the News Industry: New Business Models and Challenges (2021) by COMMUNIA Association ([invalid URL copyright text and data mining ON COMMUNIA communia-association.org])
84. Copyright in a Digital Age (Third Edition) (2022) by Liisa Hyvönen and Christopher Millard, Oxford University Press

85. The Future of Copyright: Balancing Creativity and Digital Innovation (2020) by Pamela Samuelson, Oxford University Press
86. Machine-Readable Copyright Information: A Study on Legal Frameworks and Technical Standards (2023) by World Intellectual Property Organization (WIPO)
87. The Text and Data Mining Exception and the Opt-Out Mechanism under the EU Copyright Directive (2021) by Eleftheria Marina Moschidou, European Intellectual Property Review, Vol. 53, No. 11, pp. 721-732
88. Moschidou, E. M. (2021). The Text and Data Mining Exception and the Opt-Out Mechanism under the EU Copyright Directive. European Intellectual Property Review, 53(11), 721-732.
89. The Text and Data Mining Exception and the Monetization of Text and Data (2021). Eleftheria Marina Moschidou,European Intellectual Property Review, Vol. 53, No. 3, pp. 161-173.
90. Bypassing Paywalls: Access to Scholarly Literature and the Role of Shadow Libraries (2020). Lindsay Chad & Tracey P.Lau Librarian Issue Vol. 121, No. 1
91. Challenges in Implementing the Text and Data Mining Exception in the EU by U Putra & H Liu (2021)
92. Text and Data Mining in the Context of Scholarly Publishing (2020) by Laura Quilter, Journal of Librarianship and Scholarly Communication, Vol. 8, No. 1
93. The Impact of the EU Copyright Directive on Text and Data Mining (2020) by A. Liogier
94. Legal and Technical Interoperability Challenges for Text and Data Mining (2019) by S. Ruegge
95. Challenges and Opportunities for Text and Data Mining in the Scholarly Ecosystem by The Association of American Universities (AAU) https://www.aau.edu/research/featured-research
96. The Future of Fair Use in the Digital Age by The Center for Democracy & Technology
97. The Text and Data Mining Exception and Its Impact on Competition by Max Planck Institute for Innovation and Competition https://www.science.org/doi/10.1126/science.add6124
98. A Global Approach to Text and Data Mining by The National Academies Press https://www.science.org/doi/10.1126/science.add6124

99. Can Text and Data Mining Thrive in the Shadow of Opt-Out? by University of Amsterdam (https://aihr.uva.nl/about-aihr/research-data-management/during-the-research.html)
100. The Impact of Text and Data Mining on Innovation by Center for Information Policy Leadership (CIPL) at Hunton Andrews Kurth LLP (https://www.huntonak.com/privacy-and-information-security-law/category/centre-for-information-policy-leadership)
101. Towards a Standardized Opt-Out Mechanism for Text and Data Mining in the EU by Max Planck Institute for Innovation and Competition
102. The Need for Standardized Opt-Out Mechanisms in Text and Data Mining by European Digital Rights (EDRi)
103. The Challenges of Global Copyright Harmonization in the Digital Age by University of California, Los Angeles (UCLA)
104. Hugenholtz, P Bernt (2020) 'Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?' SSRN [online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4134651 [Accessed 19 May 2024]
105. European Parliament, Directorate General for Internal Policies (2018) 'The Exception for Text and data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects' [online] Available at: https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA%282018%29604941_EN.pdf [Accessed 19 May 2024]
106. LIBER Europe (2020) 'A Copyright Exception for Text and Data Mining' [online] Available at: https://libereurope.eu/document/a-copyright-exception-for-text-and-data-mining-2/ [Accessed 19 May 2024]
107. University of Turku (2023) 'Text and data mining (TDM) in the EU: What you need to know about copyright law and data analysis' [online] Available at: https://utuguides.fi/researchdata/instructions [Accessed 19 May 2024] **(**
108. The Copyright Alliance (n.d.) 'Copyright and Journalists' [online] Available at: https://copyrightalliance.org/ [Accessed 19 May 2024]