

THE 'BLACK BOX' OF AI

Reconciling Reproduction
with the Originality Standard
in EU Copyright Law

TILBURG UNIVERSITY 7500 WORDS

KLIMENTINA
MALEEVSKA



Table of Contents

Introduction	3
CHAPTER I	4
I. <i>Introduction</i>	5
II. <i>Literature review</i>	6
III. <i>Research question</i>	9
IV. <i>Concepts</i>	9
V. <i>Scientific & Social Relevance & Methodology</i>	11
CHAPTER II	12
I. <i>Introduction</i>	13
II. <i>Doctrinal Foundations of the Originality Standard</i>	14
III. <i>The Protection Threshold for Partial Reproductions</i>	16
IV. <i>TDM in the EU Copyright Framework</i>	16
V. <i>Conclusion</i>	18
CHAPTER III	19
I. <i>Introduction</i>	20
II. <i>What is ‘Knowledge Extraction’?</i>	20
III. <i>A Reproduction as a ‘Work’</i>	21
IV. <i>Originality and AI Outputs</i>	23
V. <i>Conclusion</i>	25
CHAPTER IV	26
I. <i>Introduction</i>	27
II. <i>TDM Exception</i>	28
III. <i>Economic Perspectives</i>	30
IV. <i>Conclusion</i>	32
CHAPTER V	33
I. <i>Subquestions Answers</i>	34
II. <i>Main Question Answers</i>	36
III. <i>Final remarks and limitations to research</i>	38
Bibliography	34

CHAPTER I

Originality, Authorship, and Challenges

I. Introduction

Generative Artificial Intelligence (GenAI) is a machine-based system structured to operate through its algorithm trained through Text and Data Mining (TDM), to present outputs in response to inputs, acting with varying levels of autonomy.¹ The process of training the GenAI model requires vast amounts of data, often consisting of copyrighted works.² Trained models proficiently mimic human creativity to a certain degree, owing to the characteristics of the source materials used during training.³ Within the EU, an increasingly pertinent, yet comparatively unexplored issue is whether, and to what extent, copyright exceptions apply to outputs by GenAI.⁴

TDM is governed by the CDSMD and defined by Art.2(2) as an essential technique in the AI development stage.⁵ The CSMD introduced two mandatory exceptions for TDM activities in the face of Articles 3 & 4.⁶ The purpose of this is to permit reproductions and extraction of fully accessible works conditionally for TDM purposes, otherwise restricted by sui generis rights.⁷ Developers who wish to retain a better 'fed' GenAI model would ideally strive to fulfill the exceptions requirements, given the copyright-restricted field of mining.⁸ The introduction of these exceptions was a clearer legal field for research.⁹ However, the fact of their adoption in 2019 had brought discussion on whether these provisions had the GenAI development when drafting.¹⁰ In the hope of settling it, the soon-to-be-published EU AI Act

¹Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 2.

²Tim W Dornis, 'Generative AI, Reproductions Inside the Model, and the Making Available to the Public' (2025) *IIC* (forthcoming) <https://doi.org/10.1007/s40319-025-01582-9>.

³Senftleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 2.

⁴Quintais JP, 'Generative AI, Copyright and the AI Act' [2025] 56 - Computer Law & Security Review 106107, 4.

⁵Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 7.

⁶Christophe Geiger and Vincenzo Iaia, 'The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI' [2024] 52(1) Computer Law & Security Review: The International Journal of Technology Law and Practice, 7.

⁷R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337, 5.

⁸'The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the Failed Text and Data Mining Limitations in the CSMD Directive' (2021) *PIJIP/TLS Research Paper Series* No 66

⁹R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337, 9.

¹⁰Margoni T, 'TDM and Generative AI: Lawful access and opt-outs' [2025] SSRN Electronic Journal, 11.

explicitly links the CDSMD's TDM exceptions to the development purposes of GenAI, for functional training purposes.¹¹

In examining the full range of whether the training and its output constitute an infringement of a literary work, it is crucial to start from the roots of copyright legislative incentives. Notably, the Berne Convention Article 9(1) mandates the authors' works to have exclusive right of authorising reproduction, such as outputs of their creation.¹² A foundational principle for the EU, alongside continuous guidance from the CJEU, from all standards set by the convention.¹³ While the limitations prescribed by it for a temporary reproduction are present where it is made for the sole purpose of creating a work perceptible, or is a part of a technical process which is identical to the transmission of the work, or the reproduction itself is permitted by the law of the rights holder.¹⁴ However, the specific requirements behind this convention are the exact elements that will be examined by this paper, taking an in-depth approach to whether the reproductions of GenAI amount to an infringement.

Under EU copyright law, aligning with the international standard, the right of reproduction is construed broadly.¹⁵ Article 2 of the InfoSoc Directive further re-establishes this by defining it in a 'direct or indirect, temporary or permanent' or 'in any form, in whole or part'.¹⁶ The CJEU's establishing copyright standard of Infopaq confirms this, by stating that reproduction in part occurs when a part of a work that is the author's intellectual creation is copied.¹⁷ The specific provisions for TDM exceptions are indicative of the fact that TDM activities are considered to involve copyright-relevant acts requiring authorisation.¹⁸ It is confirmed by numerous sources that it is relevant to AI outputs.¹⁹

¹¹Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 1.

¹²Berne Convention for the Protection of Literary and Artistic Works (signed 9 September 1886, entered into force 5 December 1887) 828 UNTS 221, art 9(1).

¹³PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 4.

¹⁴Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, entered into force 5 December 1887, revised at Paris 24 July 1971, and amended 28 September 1979) 1161 UNTS 3, art 9(1).

¹⁵Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 2.

¹⁶Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10, art 2.

¹⁷Rosati E, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity' (2019) 27 *Asia Pacific Law Review* 198, 11.

¹⁸Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 2.

¹⁹DUSOLIER IES (PRIA Centre for Regulation of the Creative Economy 2025) working paper, 6.

With the 2020 ‘Shaping Europe’s Digital Future’, the EC aimed to establish the EU’s digital strategy.²⁰ Including modernising copyright rules for the CDSMD, while encouraging data and licensing solutions, was sought to be achieved by the changing digital environment.²¹ However, the existing copyright framework has been deemed insufficient to provide adequate legal certainty for the advancing domain of GenAI research.²² The following literature review will examine whether and to what extent the EU copyright framework is reflecting the EU’s strategy for digital development through its legislation.²³

II. Literature review

TDMs are used purely for computational analysis, rather than employing the work’s expressive content.²⁴ Thus, it does not correspond to the typical ‘use’ or ‘exploitation’ of work of Article 9(1) of the Berne Convention.²⁵ Where the concept of a reproduction was tied to the incentive that the reproduction would be sufficiently stable to be communicated to individuals. Meanwhile, a GenAI model does not use the data fed to it in its original form, which is perceptible to an individual. Therefore, it does not trigger the reproduction definition of the Articles prescription. In this view, Senftleben emphasised that a TDM constitutes a ‘new category of copyright’ that does not exist in the classic context of purely transient copyright and that lies beyond the scope of the Article.²⁶ Given the historical absence of GenAI and their associated training methodologies at the time of the completion of these articles, it is reasonable to assert that a TDM approach is not included within their purview.

Many arguments have been posed against the TDM exceptions, mainly in light of a reasonable setback and disadvantage of the EU AI sector.²⁷ Citing one of many is Hugeholtz’s view, which prescribes it as unsuitable and far too restrictive, as interpreted in the EU’s overregulation and perspective.²⁸ The definition employed in the CDSMD is indeed ‘future

²⁰ European Commission, *Shaping Europe’s Digital Future* (COM(2020) 67 final, 19 February 2020).

²¹ Margoni T, ‘TDM and Generative AI: Lawful access and opt-outs’ [2025] SSRN Electronic Journal, 8.

²² Ibid

²³ Kretschmer M and Margoni T, ‘A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology’ [2021] SSRN Electronic Journal, 21.

²⁴ Ibid

²⁵ Berne Convention for the Protection of Literary and Artistic Works (adopted 9 September 1886, entered into force 5 December 1887, revised at Paris 24 July 1971, and amended 28 September 1979) 1161 UNTS 3, art 9(1).

²⁶ Ibid

²⁷ Geiger C, ‘The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the Failed Text and Data Mining Limitations in the CSDM Directive’ [2021] SSRN Electronic Journal, 6.

²⁸ Michael W. Carroll, ‘Copyright and the Progress of Science: Why Text and Data Mining Is Lawful’ (2019) 53 UC Davis L

proof', it aligns with the definition of most AI/ML existing and will exist as long as data analytics are considered.²⁹ Placing a narrow interpretation of exceptions, within an outdated regulating system, puts the European development plans in question.³⁰ Furthermore, the CDSMD Directive itself does not explicitly delineate the mining of the subject matter, which is not safeguarded by either sui generis rights or copyright.³¹ An additional area of ambiguity not addressed by the CDSMD Directive pertains to whether a researcher within the EU may transfer a lawfully acquired database to a research partner in the Union for data mining. This raises questions regarding the concepts of 'lawful access' and 'lawful use' as outlined in the temporary reproduction exception, which is unclear.³² Addressing the incentives targeted in formulating the CDSMD Directive is essential, which supports the framework established by the Marrakesh Treaty. This consideration necessitates reflecting on the underlying, and still widely accepted, assumption that only exclusive rights can facilitate innovation and creativity.³³

There are two notably differing scholarly positions on the topic. The first one is against copyright protection for autonomous AI outputs, where it is accepted that generated outputs, particularly with minimal human intervention, should not receive copyright protection.³⁴ The core rationale being that traditional copyright law is anthropocentric, requiring a human author.³⁵ With concepts like the Lockean labour theory or Kantian personality lacking consciousness, emotion, or intention in the stages of the creative process.³⁶ Deeming the consequential output as lacking requisite originality for protection.³⁷ While the second view is in favour of copyright protection. Basing its position on the fact that many of the outputs do involve personalised traits by the inputs of the consumer, personalising it and creating a

Rev 893, 6.

²⁹Kretschmer M and Margoni T, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' [2021] SSRN Electronic Journal, 3.

³⁰ Federico Ferri, 'The Dark Side(s) of the EU Directive on Copyright and Related Rights in the Digital Single Market' (2020) China-EU L J <https://doi.org/10.1007/s12689-020-00089-5>.

³¹ *ibid*

³²Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' (Centre for International Intellectual Property Studies (CEIPI) 2019) 7 <https://papers.ssrn.com/abstract=3470653> accessed 26 February 2020; Maria Bottis and others, 'Text and Data Mining in the EU Acquis Communautaire Tinkering with TDM & Digital Legal Deposit' (2019) 12 Erasmus L Rev 179, 179.

³³ S Flynn and others, 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action' (SSRN, 2020) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3578819 accessed 10 March 2025.

³⁴Saw CL and Lim D, 'The Case for AI Authorship in Copyright Law' [2025] SSRN Electronic Journal, 1.

³⁵Bonadio E and McDonagh L, 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity' (Sweet and Maxwell 2020), 2.

³⁶Nathan Pelgrims, 'Lockean theories of intellectual property and the product of artificial intelligence' (Master's thesis, Universiteit Gent 2021), 2.

³⁷Ramalho A, 'Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems' [2017] SSRN Electronic Journal, 13.

jointly authored work.³⁸ While the primary justification being from a utilitarian perspective, is to incentivise investment and development in creative technologies.³⁹ Reflecting Recital 5 of the InfoSoc Directive calls for adapting copyright law for technological change, acknowledging the need to accommodate new forms of creative processes.⁴⁰ Confirming a view in support of legislative reforms of already existing standards.

III. Research question

The previous discussion led to the following main research question: *How can the CJEU's approach to originality in partial copying be applied to generative AI's latent knowledge extraction, and what gaps does this reveal in the EU copyright framework on originality?*

The research question shall be answered through the following subquestions:

- *How has EU case law defined originality and the protection threshold for partial reproductions, and how does this standard apply doctrinally to generative AI?*
- *What is 'latent knowledge extraction' in generative AI, and when might AI-generated outputs amount to partial copying of protected works?*
- *To what extent do current EU copyright laws effectively regulate both the reproduction of protected content during AI training, and what doctrinal gaps or policy tensions does this dual analysis reveal for potential regulatory reform?*

IV. Concepts

A. Originality

³⁸Senftleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 9.

³⁹Bonadio E and McDonagh L, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Sweet and Maxwell 2020) , 12.

⁴⁰Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10, recital 5.

Originality is a fundamental copyright principle that protects the form of the expression.⁴¹ Defining AI reproduction copyright boundaries is now clearly established by Infopaq and subsequent case law.⁴² This standard has shaped the phrasing behind, as well as the recommended standards for classifying subsequent legislation regarding AI, which have formed its roots beforehand. Assessing whether the definition of originality aligns with the expected notion of reproduction is crucial.

B. Text and Data Mining

TDM is an automated analytical technique used to extract information, patterns, trends, and correlations from large quantities of digital text and data.⁴³ This process involves using computational methods to analyse content, which can include journal articles, books, websites, social media, images, and audio files.⁴⁴

C. Partial Reproduction

The reproductions generated by AI models as a result of TDM training are the key area of examination in the infringement cases, making them a mandatory subject of discussion within this thesis.⁴⁵ Essentially, it assesses its originality in relation to its technological context.

V. Scientific & Social Relevance

This topic offers significant scientific novelty as this research critically analyses existing AI definitions within scientific literature and legal frameworks to assess their relevance, comprehensiveness, and topicality. The aim is to formulate a new definition that encompasses all significant AI properties, thereby contributing to ongoing scientific discussions on the multifaceted concept of AI and developing a more suitable conceptual framework, as well as with ongoing cases, such as *Like Company v Google Ireland*.⁴⁶

⁴¹Lemley M, 'How Generative Ai Turns Copyright Law Upside Down' (2024) 25 Science and Technology Law Review, 6.

⁴²Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465.

⁴³M. Barnwell, 'Balancing the benefits of TDM against copyright protection: A comparative study on the exceptions and limitations to the reproduction right with regard to text and data mining in the EU and the US' (LLM Thesis, Tilburg University 2018), 20.

⁴⁴Svensson G 'Text and Data Mining in EU Copyright Law' (Uppsala University, 2020), 8.

⁴⁵Senftleben M, 'Compliance of National TDM Rules with International Copyright Law – an Overrated Nonissue?' [2022] SSRN Electronic Journal, 7.

⁴⁶F. Ceci, F. Masciarelli and S. Poledrini, 'How social capital affects innovation in a cultural network: Exploring the role of bonding and bridging social capital' (2019) 17 Eur J Innovation Management 973, 17.

This research holds practical significance because of the current inability to apply existing legal norms to AI, given the lack of specific legal definitions and rules governing its use.⁴⁷ The findings, particularly in defining AI, can enhance legal regulation mechanisms. Furthermore, existing anthropocentric legal models are insufficient to integrate AI into societal relations. Therefore, this investigation is crucial for maintaining social system stability amid widespread AI implementation, and it supports the need for a new concept of a legal subject as AI continues to integrate further into social practices.

Methodology

The methodological approach employed to analyse the research questions and critically assess the EU Copyright Framework concerning AI use and its outputs is doctrinal legal analysis. This choice gives the explanatory, axiomatic, and logical nature of the arguments.⁴⁸ It is ideally suited for examining doctrines within the European Union's copyright framework. This method's strength lies in its ability to clarify legal principles as a discipline and illuminate the nuanced application of law within the evolving context of AI. Furthermore, this thesis critically and reflectively scrutinises the coherence, effectiveness, and future implications of current EU copyright provisions in light of technological advancements in AI.

VII. Chapter Overview

This thesis is divided into four subsequent chapters, which answer a sub-question and advance the main investigation into whether EU copyright legislation sufficiently governs the extraction of latent information by generative AI. The theoretical foundations are laid out in Chapter II through an analysis of the EU's originality requirement and the notion of partial replication, as informed by its legal precedents. Following Chapter 3, examining the technical aspects of a reproduction and how the framework governs such an output by GenAI in the copyright sphere. Chapter IV will determine whether the current policies align with the EU's strategy. Followed by Chapter V, answering the main research question and concluding a critical analysis of the legislative gaps.

⁴⁷I.R. Begishev, E.Y. Latypova and D.V. Kirpichnikov, 'Artificial Intelligence as a Legal Category: Doctrinal Approach to Formulating a Definition' (2020) 14 Actual Problems of Economics and Law 79-91, 3.

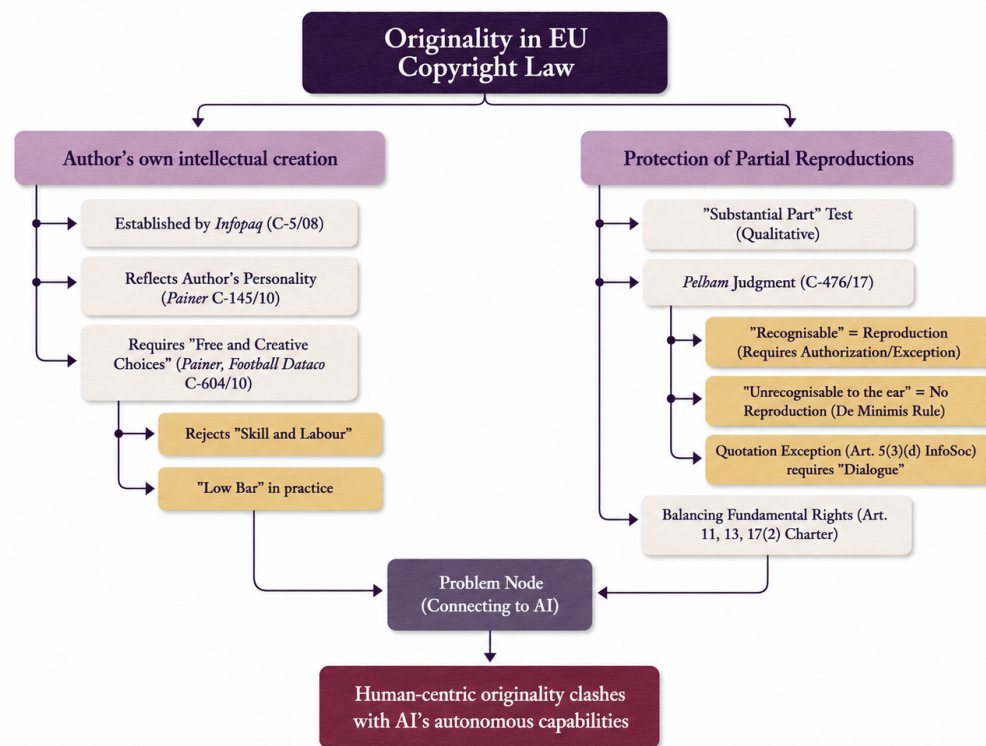
⁴⁸I.J. Kroeze, 'Legal Research Methodology and the Dream of Interdisciplinarity' (2013) 16(3) PER, 13.

CHAPTER II

Originality Standard and Consequences to TDM Training

I. Introduction

EU copyright's originality standard dictates the creator-work relationship, focusing on unique creative expression over effort or time. Rooted in the CJEU's landmark *Infopaq* judgment, this standard centres on 'the author's intellectual creation'.¹ CJEU precedents consistently affirm that this 'intellectual creation' must stem from a natural person's 'free and creative choices,' reflecting their personality.² This chapter will analyse how EU case law has defined originality and the protection threshold for partial reproductions, and critically examine the doctrinal application of these standards to generative AI contexts by answering the question: *How has EU case law defined originality and the protection threshold for partial reproductions, and how does this standard apply doctrinally to generative AI?*



¹Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para.49.

²*Football Dataco Ltd v Sportradar GmbH* (C-604/10) EU:C:2012:300, para.40.

II. Doctrinal Foundations of the Originality Standard

The EU's originality standard, definitively established in the *Infopaq* case, governs copyright protection requirements, emphasising the creator's creativity, autonomy, and human authorship.³ This requires a qualitative assessment rather than a quantitative measure based on effort or 'skill and labour'.⁴ The standard depicts the need for autonomous creative decision-making without defining source limits, which inherently restricts authorial autonomy.⁵ Therefore, it decimates the need for the autonomous decision-making of the creator as a descriptive component of the evaluation.⁶

Google and other AI developers commonly claim that GenAI outputs are commonly the product of hallucination.⁷ It has been demonstrated that generative AI systems may produce believable but inaccurate or non-factual information, a condition known as hallucination.⁸ This defence argues that AI 'predicts' content based on probabilities, rather than explicit recreation, aiming to disprove 'copying' as a basis for infringement. EU copyright law does not expressly define 'authorship' for AI-generated content, instead relying on a human-centric interpretation of the Berne Convention and CJEU jurisprudence, which consistently presupposes a human author. A key doctrinal point: while AI can assist creativity, copyright for AI-assisted works hinges on human authors exercising 'sufficient creative control.'⁹ This means the human must exercise 'free and creative choices' to 'stamp their personal touch' on the work, rather than the AI acting autonomously.¹⁰ Scholars support this, emphasising implicit human authorship where authors conceive the work before expression.¹¹

The CJEU emphasises that subjective value judgements are not used to evaluate originality, and it does not require creativity or personality to be objectively discernible in terms of

³Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para. 58.

⁴PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 7.

⁵S van Gompel and M van Eechoud, 'Creativity, Autonomy and Personal Touch: A Critical Appraisal of the CJEU's Originality Test for Copyright' in M van Eechoud (ed), *The Work of Authorship* (Amsterdam University Press 2014), 96.

⁶*ibid*

⁷McDaniels Law, 'First Case On Whether AI Fits Into EU Copyright Law Referred To The CJEU' (McDaniels Law, 2024) <https://mcdanielslaw.com/first-case-on-whether-ai-fits-into-eu-copyright-law-referred-to-the-cjeu/> accessed 5 May 2025.

⁸William Fry, 'CJEU to Rule on AI Chatbots and Copyright in Landmark Case Against Google' (William Fry, 2024) <https://www.williamfry.com/knowledge/cjeu-to-rule-on-ai-chatbots-and-copyright-in-landmark-case-against-google/> accessed 1 June 2025.

⁹PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 12.

¹⁰D.L. Burk, 'Thirty-Six Views of Copyright Authorship, by Jackson Pollock' (2020) *J World Intell Prop*, 320.

¹¹Bonadio E and McDonagh L, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Sweet and Maxwell 2020), 3.

aesthetic merit or quality.¹² While low, the originality bar is not non-existent. Purely functional works, dictated by technical constraints, or so trivial or banal that they leave no room for creative freedom, are unlikely to qualify for protection. Aligning with the principle that intellectual effort and skill alone are insufficient if they do not embody any creative choices, as reiterated in cases like *Football Dataco* and *BSA*.¹³ Indeed, *Painer* further solidified that the originality standard does not depend on the *degree* of creativity, reinforcing the 'low bar' for protection, provided the author's personality is reflected through their choices.¹⁴ Painer's analysis posits the question of the range of creativity needed to be exhibited for a work to be defined as original.¹⁵ This flexibility allows for originality to be demonstrated at various stages of creation, including through personal input that modifies or arranges AI-generated content, potentially leading to the delineation of copyrightable elements within an AI-assisted work. Further, it clarified that an intellectual creation is considered the author's own if it reflects their personality.¹⁶ This means that the author must have been able to 'stamp his/her personal touch on the work created' through their choices in preparation, execution, and development.¹⁷

Despite clarifications, the CJEU hasn't provided explicit guidance on assessing the creativity of AI-assisted work or a minimum creative input threshold for AI development. Authors must show creative self-expression in work creation. This need, on the other hand, is not satisfied when the job is determined by technical considerations, rules, or constraints which leave no room for creative freedom.¹⁸ The CJEU emphasised the need for free and creative choices and a personal touch for uniqueness, especially when it comes to database creation, as seen in *Football Dataco*. That said, the CJEU has not attempted to give further guidance on the topic of assessing 'creativity', even less defining a minimum threshold for creative manners within the AI development field.¹⁹ A related, yet insufficiently examined, dimension concerns the originality threshold for creations mandated by technical considerations that leave no room for creative freedom, as established in *Bezpečnostní softwarová asociace*.²⁰ Whereas, in

¹²PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 42.

¹³*Football Dataco Ltd v Yahoo! UK Ltd* [2012] EWHC 1400 (Ch); *Football Dataco Ltd v Yahoo! UK Ltd* [2013] EWCA Civ 1583; *Football Dataco Ltd v Sportradar GmbH* (C-604/10) EU:C:2012:300; *BSA v Ministry of Defence* [1991] EWHC Ch 1.

¹⁴Case C-145/10 *Painer v Standard VerlagsGmbH and Others*, Opinion of AG Trstenjak, ECLI:EU:C:2011:798 119-120.

¹⁵*ibid.* 124.

¹⁶*ibid.* 119-120.

¹⁷ *ibid.* 124.

¹⁸*Football Dataco and Others v Yahoo! UK Ltd and Others* EWCA Civ 1380, 2 WLR 400, 39.

¹⁹PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 9.

²⁰Case C-393/09 *Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury* [2010] ECLI:EU:C:2010:816, para. 76.

Painer, the CJEU had clarified the possibility of creating a work with an authorship nature that had the influence and aid of a machine or device.²¹ However, the CJEU emphasised free and creative choices and a personal touch for database uniqueness.²²

III. The Protection Threshold for Partial Reproduction: The ‘Substantial Part’ and ‘De Minimis Test’

Initially, the CJEU upheld the notion that, even if a sound sample is very brief, it still constitutes a reproduction of a portion of a phonogram, thereby safeguarding the producer's investment in the original work. A crucial distinction was added by the Court, nonetheless, which states that a sample's usage does not constitute a reproduction under the InfoSoc Directive if it is not audible in the new work. Setting a *de minimis* criterion for unidentifiable samples exemplifies the judiciary's practical balancing act. The Court achieved this by weighing various fundamental rights outlined in the EU Charter. This strategy demonstrates that, even in cases where a technical replication has occurred, the CJEU is willing to identify certain domains where rigorous copyright enforcement would unnecessarily restrict creative freedom. This balancing process implies that the extent of copyright protection is not absolute and may be limited by concerns over fundamental rights, leading to debates over AI outputs that are highly transformative or indistinguishable from human works.

IV. TDM in the EU Copyright Framework

TDMs use large datasets to train AI algorithms, limited to the computing process's input stage.²³ Their scope typically limits to data acquisition and analysis, not necessarily extending to trained model integration or output generation.²⁴ While the *sui generis* database right aims to protect the investment of database creators, the TDM exceptions themselves are designed to facilitate data analysis.²⁵ A significant challenge arises because the material used for TDM, particularly as input for AI models, often holds substantial economic value, may be

²¹Case C-145/10 *Painer v Standard VerlagsGmbH and Others*, Opinion of AG Trstenjak, ECLI:EU:C:2011:798, para.5.

²²*Football Dataco Ltd v Sportradar GmbH* (C-604/10) EU:C:2012:300, 35-36.

²³R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 *European Intellectual Property Review*, 2021/5, p. 322-337.

²⁴Adam Buick, ‘Copyright and AI Training Data—Transparency to the Rescue?’ (2025) 20(3) *Journal of Intellectual Property Law & Practice* 182 <https://doi.org/10.1093/jiplp/ipae102>.

²⁵Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L157/1, art 2.

confidential, or is protected as a trade secret.²⁶ Raising critical questions about how creators whose works are used in TDM can identify such use and effectively seek remuneration.

Several challenges persist due to these limitations, including a narrow scope and systematic inconsistencies in the development of these models, such as security measures, storage of works, admissibility, and the remuneration of authors, as well as the legislative incentives behind these provisions. Firstly, the TDM research exception's imprecise scope is criticised for potentially failing to accommodate scientific research realities and risking researcher penalisation.²⁷ Secondly, security obligations disparity exists: while some exceptions imply secure storage, CDSM Article 4 doesn't explicitly oblige developers to implement specific security measures for stored copies.²⁸ Raising concerns, primarily as Article 4 is often used for commercial AI training, a non-research purpose requiring careful consideration in commercial settings.²⁹ This contrasts with Article 3, which is specifically for scientific research and has different retention rules.³⁰ Third, Article 4 does not permit retaining TDM-generated copies for result validation, significantly contrasting with Article 3(2)'s explicit allowance for scientific verification.³¹ Finally, while the individual economic significance of a single work used in TDM or AI training may be minimal, the cumulative commercial exploitation of vast quantities of creative works, particularly where TDM exceptions are not applicable or are contested, creates a significant dissonance between the law's current focus and commercial incentives.³²

V. Conclusion

CJEU jurisprudence defines originality as the author's intellectual creation, necessitating creative choices of the author's individuality.³³ This criterion sets a low bar and distinguishes it from skill and labour.³⁴ Partial reproductions are protected by the qualitative substantial part

²⁶ibid

²⁷R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 *European Intellectual Property Review*, 2021/5, p. 322-337.

²⁸Directive (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, art 3(2).

²⁹ibid

³⁰Christophe Geiger, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative AI' [2024] 55(7) *IIC - International Review of Intellectual Property and Competition Law* 1129-1165, page 4.

³¹Quintais JP, 'Generative AI, Copyright and the AI Act' [2025] 56 - *Computer Law & Security Review* 106107, page 6.

³²Quintais JP, 'Generative AI, Copyright and the AI Act' [2025] 56 - *Computer Law & Security Review* 106107, page 13.

³³Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para. 58.

³⁴Synodinou, Tatiana, Book Chapter: The Foundations of the Concept of Work in European Copyright Law (January 30, 2012). Synodinou (ed.), "Codification of European Copyright, Challenges and perspectives, the Hague, Kluwer Law International, p. 93-113, 2012, 103.

test, meaning that even small portions may infringe upon IP rights. Unrecognisable sound samples don't qualify as reproduction, establishing a critical *de minimis* criterion balancing copyright and artistic expression.³⁵

³⁵*Pelham GmbH v Ralf Hütter* (C-476/17) EU:C:2019:624, para, 72-82.

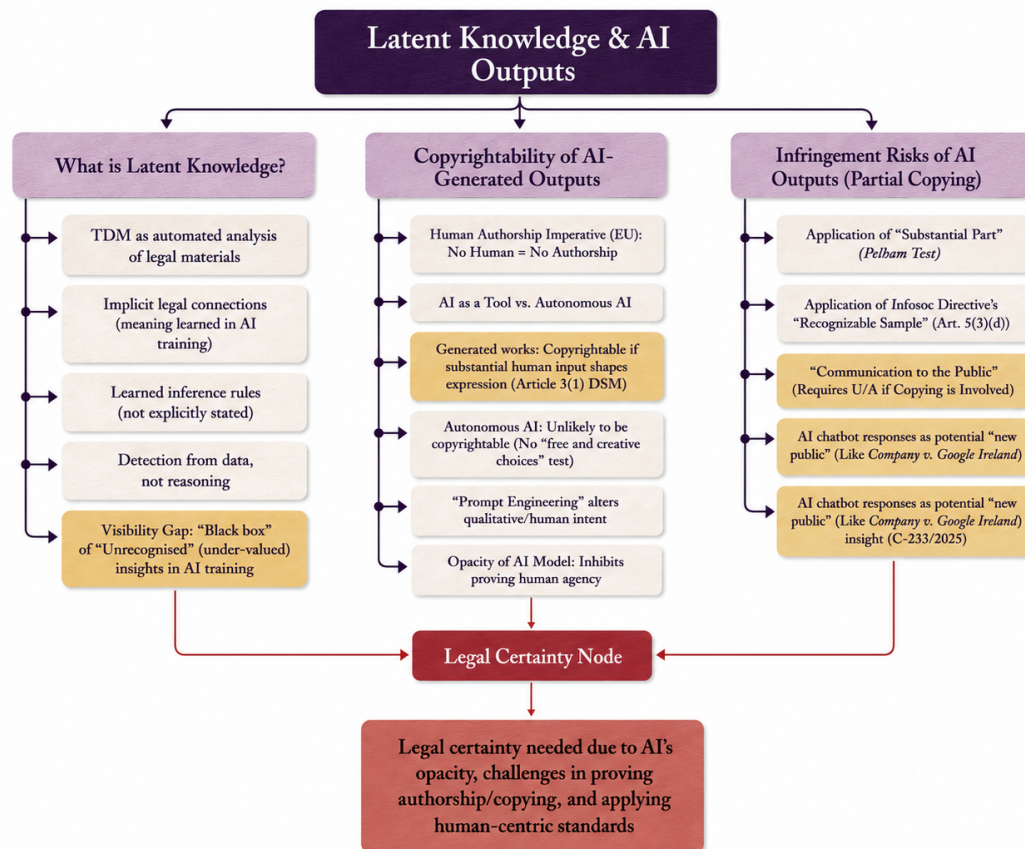


CHAPTER III

Originality, Authorship, & Infringement of AI

I. Introduction

GenAI models produce outputs reflecting patterns and knowledge from training data, raising complex questions about their relationship to the authors' creative expressions.¹ This chapter examines the technical process of 'latent knowledge extraction' in GenAI, exploring how TDM input influences AI outputs. Further analysing when these AI-generated outputs might amount to partial copying of protected works.² Specifically, this chapter aims to answer: *What is 'latent knowledge extraction' in generative AI, and when might AI-generated outputs amount to partial copying of protected works?*



II. What is 'Latent Knowledge Extraction'?

¹Feuerriegel S and others, 'Generative AI' [2023] SSRN Electronic Journal, 2.

²Geiger C, Frosio G and Bulayenko O, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' [2019] SSRN Electronic Journal, 7.

In GenAI models, particularly diffusion models, input data is converted into embedding vectors and represented in a multi-dimensional latent space. This latent space internally processes and represents information, training AI for generating outputs. TDM is an automated analytical method examining vast datasets, called knowledge discovery.³ TDM can glean new information from literature, finding new research directions or unexpected connections.⁴ While traditional TDM extraction duplicates content, in AI models, can also use the users prompting a near-identical training data replica.⁵ This differs from memorisation, where a copy recreates without user intent, or regurgitation, an unintended, direct duplication of training data.⁶ Therefore, latent knowledge extraction in GenAI refers to an AI model analysing large datasets to abstract patterns, trends, and correlations, encoding this learned information within its internal latent space.⁷

III. A Reproduction as a ‘Work’

For a GenAI output to be recognised as an original 'work' under EU copyright law, it must satisfy the criteria derived from the author's intellectual creation, which implicitly requires human authorship.⁸ Specifically, four requirements must be satisfied. First, outputs must fall within a literary, scientific, or artistic domain, achievable through training.⁹ Second, the output must tangibly connect to human input.¹⁰ While currently seemingly remote, this could be met under specific conditions. Purely autonomous outputs of content are generally ineligible for copyright protection, AI-assisted works can be protected if a human exercises sufficient creative control, utilising AI as a tool. This is satisfied when a user's prompt or modifications significantly influence the output's expressive elements, reflecting human creative influence and personal touch. Thirdly, the output must embody the author's

³Rosati E, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of Ai Creativity' (2019) 27 Asia Pacific Law Review 198, page 13.

⁴Michael W. Carroll, 'Copyright and the Progress of Science: Why Text and Data Mining Is Lawful' (2019) 53 UC Davis L Rev 893, 12.

⁵Maria Bottis, Marinos Papadopoulos, Christos Zampakolas & Paraskevi Ganatsiou, 'Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit' (2019) 12 Erasmus L Rev 190, 7.

⁶Cooper, A. Feder and Grimmelmann, James, 'The Files are in the Computer: Copyright, Memorization, and Generative AI' (April 22, 2024). Cornell Legal Studies Research Paper Forthcoming, Chicago-Kent Law Review, Forthcoming, 3.

⁷M. Barnwell, 'Balancing the benefits of TDM against copyright protection: A comparative study on the exceptions and limitations to the reproduction right with regard to text and data mining in the EU and the US' (LLM Thesis, Tilburg University 2018), 17.

⁸Christophe Geiger, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative AI' [2024] 55(7) IIC - International Review of Intellectual Property and Competition Law 1129-1165, page 9.

⁹PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 IIC 1190, page 11.

¹⁰ibid

personality.¹¹ These creative elements can manifest at various levels of the work and be introduced throughout the creative process, from conception and execution to the final arrangement or modification of the output. Notably, these creative elements can be introduced throughout the process, particularly during the stages of conception, execution, and redaction.¹² Scholarly consensus affirms that the act of creation may consist in the thoughtful configuration of pre-existing materials, thereby transforming the ‘ready-made’ into a work suitable for presentation.¹³ This notion has been invoked for CDSMD's expansive TDM exception, which recognises that sufficiently trained AI can, with user direction, replicate human-like creativity. Demand for such AI-generated expressive content further underscores this.¹⁴ Further highlighted by the growing demand for expressive content generated through such means. CJEU jurisprudence suggests that fulfilling this requirement is straightforward if it demonstrates autonomous and creative decision-making. This study primarily concentrates on the latter two stages, as they are comparatively weaker in satisfying the criteria.

While the *Infopaq* judgment primarily established the originality standard, it also affirmed the broad interpretation of the reproduction right under EU law. Therefore, defining the GenAI reproduction process is crucial for assessing whether outputs qualify as EU copyright reproductions. Notably, the term reproduction is not explicitly defined in EU copyright statutes.¹⁵ Construed widely for high protection and due compensation for rights owners.¹⁶ Furthermore, the CJEU adopts a low-threshold approach for actionable reproduction. As *Pelham I* clarified, reproduction may infringe if it (a) recognises the original work and (b) reflects the rightholder's investment.¹⁷ This approach underscores the CJEU's emphasis on protecting both the economic and moral interests of authors.¹⁸ Additionally, platforms purporting to respect third-party rights often fail to ensure compliance. These considerations underscore the persistent legal ambiguities surrounding the copyright status of activities such

¹¹Case C-145/10 *Painer v Standard VerlagsGmbH and Others* [2011] ECLI:EU:C:2011:798, paras 87–88.

¹²*ibid*

¹³AIPPI German Delegation, 'Copyright in Artificially Generated Works' (National Report, 2019) <https://aippi.soutron.net/Portal/Default/en-GB/RecordView/Index/254> accessed 15 April 2021.

¹⁴Senfleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 13.

¹⁵Geiger C, Frosio G and Bulayenko O, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' [2019] SSRN Electronic Journal, 9.

¹⁶Margoni T, 'Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns Ai?' [2018] SSRN Electronic Journal, 14.

¹⁷*Pelham GmbH and Others v Ralf Hütter and Florian Schneider-Esleben* (C-476/17) EU:C:2019:624, [33], [35], [37], [74], [97].

¹⁸*Frank Peterson v Google LLC and YouTube LLC* (Case C-682/18) EU:C:2021:503.

as memorisation at the model training stage and reconstruction during output generation, which remain largely unresolved under current EU law.¹⁹

The CJEU requires reproduction to be recognisable and demonstrate an 'intention of entering into dialogue' with the original work.²⁰ This suggests source traceability, a significant challenge for AI. However, this same traceability, according to Jeroen Zweer, is described as almost impossible and a guessing game.²¹ The works used to train GenAI models are internalised abstractly as digital information in parameter networks, not as direct duplicates.²² Under the *Infopaq* framework for reproduction infringement, a plaintiff must first prove the work's protection and then demonstrate that the defendant copied a protected part of it. This becomes exceedingly difficult when the original work is not directly replicated but rather abstractly represented within the model.²³ Consequently, the human input requirement, a cornerstone of EU copyright doctrine, becomes practically challenging to enforce due to the technical impossibility of tracing specific human creative contributions within AI-generated outputs. Zweer's casino analogy perfectly captures the unpredictability and lack of direct control over the output's lineage.²⁴ This reveals a critical practical gap in the application of the originality standard, moving beyond a purely doctrinal discussion.

IV. Originality and AI Outputs

Copyright law's rejection of algorithmic authorship is rooted in its anthropocentric view.²⁵ However, research indicates that specific GenAI models, during output generation, can employ memorisation techniques, leading to outputs that mimic, continue, or directly

¹⁹European Copyright Society, 'Copyright and Generative AI: Opinion of the European Copyright Society' (January 2025) <https://ssrn.com/abstract=0000000>, 7.

²⁰Rosati E, 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' [2024] European Journal of Risk Regulation 1-25, 2.

²¹Interview with Jeroen Zweers (Tilburg, 22 May 2025).

²²J.B. Nordemann and J. Pukas, 'Copyright Exceptions for AI Training Data—Will There Be an International Level Playing Field?' (2022) 17(12) JIPLAP 973, 2.

²³TimW Dornis, 'Generative AI, Reproductions Inside the Model, and the Making Available to the Public' [2025] 56(6) IIC - International Review of Intellectual Property and Competition Law 1-30, 5.

²⁴Interview with Jeroen Zweers (Tilburg, 22 May 2025).

²⁵Christophe Geiger and Vincenzo Iaia, 'The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI' [2024] 52(1) Computer Law & Security Review: The International Journal of Technology Law and Practice, 5.

reproduce patterns and content from their training inputs.²⁶ This increases with model size, answer repetition, and prompting methods.²⁷

A primary challenge with GenAI models stems from their heuristic and often opaque 'black box' nature. Unlike rule-based systems, GenAI relies on probabilistic patterns learned from vast training data, rather than rigid, method-based responses.²⁸ This inherent unpredictability, coupled with phenomena like hallucinations, makes it difficult to trace the precise lineage of outputs and identify specific sources.²⁹ This opacity directly complicates copyright application, obscuring human authors' free and creative choices and infringing on the origin of the content.³⁰

Second, to prove model infringement, a plaintiff must evidence that training enabled copying exceeding a de minimis amount of protectable expression.³¹ Legislation specifies that models should store works used for the process, which are not available in the model itself.³² Furthermore, the plaintiff would have to prove his work itself was used, bringing an additional complication given that the training material might be perceived as a trade secret.

Additionally, a rightsholder could only sue in the state where the training occurred.³³ Incentivising developers to choose advantageous training locations strategically. However, the EU AI Act aims to mitigate such jurisdiction shopping by mandating that all general-purpose AI models offered in the EU must comply with EU copyright laws, including enabling content owners to opt out from TDM.³⁴ This extraterritorial reach means non-EU developed models must adhere to these transparency and opt-out requirements if placed on the EU market.³⁵ Presenting significant technical and logistical challenges for AI developers, as training content is often adjusted iteratively, and 'ex ante' reservations or opt-outs would need to be integrated into a dynamic development process. Moreover, EU AI Act Recital 105

²⁶Mezei P, 'A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI' [2024] SSRN Electronic Journal, 5.

²⁷ibid. (2).

²⁸Margoni T and Kretschmer M, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685, 4.

²⁹Gwagwa A (The Royal Institute of International Affairs Chatham House 2024) research paper, 54.

³⁰Drexler J and others, 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Perspective' [2019] SSRN Electronic Journal, 7.

³¹Lemley M, 'How Generative AI Turns Copyright Law Upside Down' (2024) 25 Science and Technology Law Review, 14.

³²Margoni T and Kretschmer M, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685, 13.

³³Peukert, 'Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training', [2024] Research Paper of the Faculty of Law of Goethe University Frankfurt/M. No. 3/2024, 26.

³⁴David Bomhard und Jonas Siglmüller, 'AI Act – das Trilogieergebnis' (2024) Recht Digital, 45-46.

³⁵Nestor Maslej and others, *The AI Index 2025 Annual Report* (Stanford University, Institute for Human-Centered AI, April 2025) <https://ai100.stanford.edu/> accessed [10 April 2025].

states that copyrighted TDM, a reproduction, is lawful only with permission or under exceptions.³⁶ Consequently, the training phase, not strictly transient or authorised, would not meet InfoSoc Directive Article 5(1) and would fall under reproduction rights.³⁷

To demonstrate the inherent difficulty of establishing infringement of AI outputs produced by ‘latent knowledge extractions’, the study identifies several interrelated reasons why pinpointing individual copyrighted sources used during the training process is challenging. The method involves a substantial amount of content. Much content is copyright-eligible due to low protection thresholds, increasing volume and complexity.³⁸ Third, there is no necessary register for copyrighted works. Finally, the projected percentage of orphan works that may be copyright works is above 40%, indicating a poor condition of rights ownership for data.³⁹

V. Conclusion

‘Latent knowledge extraction’ training refers to the automated process where AI algorithms analyse large datasets to identify, learn, and abstract patterns, trends, and correlations, encoding this knowledge within the model's internal architecture.⁴⁰ The model's architecture further encodes this learned data, comprising the resulting reproduction.⁴¹ This is distinct from simply extracting or copying data for input or storage purposes. However, the pattern recognition is part of the training phase. The output phase occasionally results in the replication of training data content, which may give rise to copyright issues.⁴²

³⁶Quintais JP, ‘Generative AI, Copyright and the AI Act’ [2025] 56 - Computer Law & Security Review 106107, 5.

³⁷Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10, art 5(1).

³⁸Buick A, ‘Copyright and AI Training Data—Transparency to the Rescue?’ (2024) 20 Journal of Intellectual Property Law and Practice 182, 10.

³⁹Rosati E, ‘Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of Ai Creativity’ (2019) 27 Asia Pacific Law Review 198, 6.

⁴⁰Guadamuz A, ‘A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs’ [2023] SSRN Electronic Journal, 26.

⁴¹Rosati E, ‘Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law’ [2024] European Journal of Risk Regulation 1-25, 7.

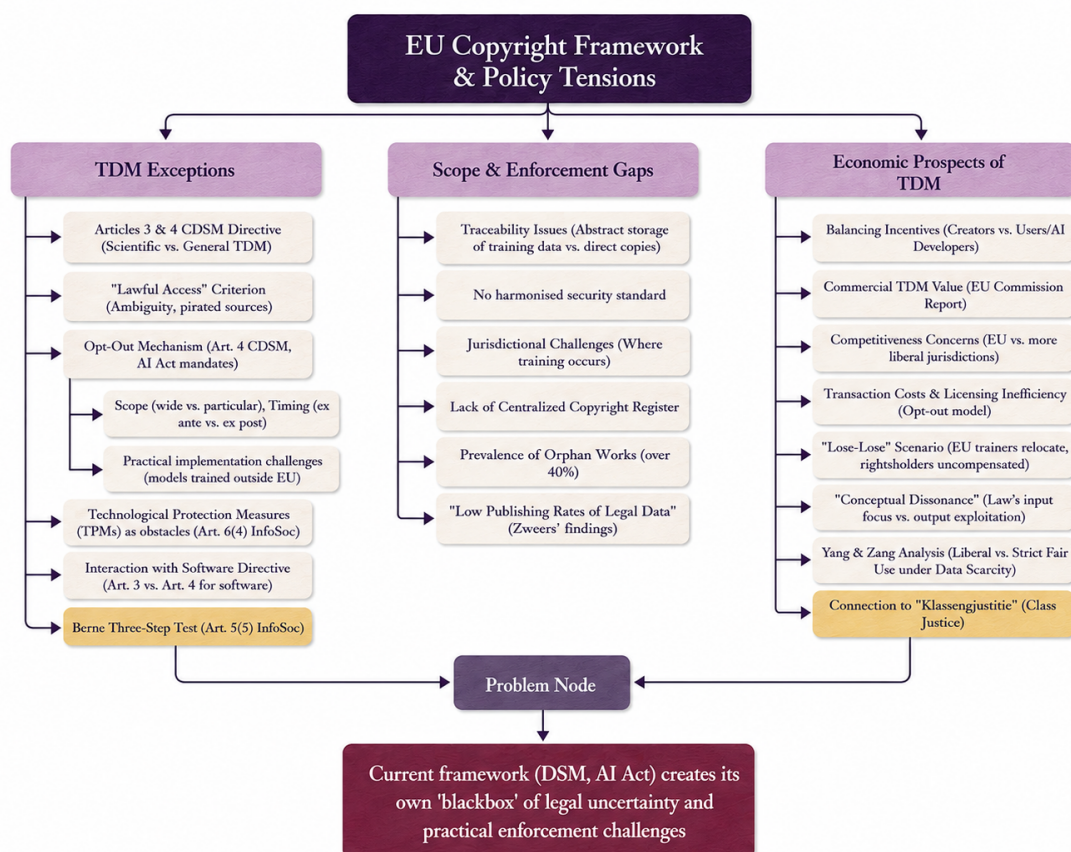
⁴²Bonadio E and McDonagh L, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Sweet and Maxwell 2020), 4.

CHAPTER IV

Regulatory Coherence, Policy Tensions, and Reform Proposals

I. Introduction

Significant legal ambiguities persist concerning the interaction between the DSM Directive, the AI Act, and the technical realities of generative AI reproduction.¹ This is mainly because the TDM exceptions were formulated before the full scope and nature of generative AI research could be predicted.² Therefore, this chapter aims to answer the question: *To what extent do current EU copyright laws effectively regulate both the reproduction of protected content during AI training, and what doctrinal gaps or policy tensions does this dual analysis reveal for potential regulatory reform?*



II. TDM Exceptions

¹Senftleben M, 'Compliance of National TDM Rules with International Copyright Law – an Overrated Nonissue?' [2022] SSRN Electronic Journal, 25.

²Senftleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 8.

Despite the defined scope of TDM, it remains unclear whether it encompasses all copyright-relevant activities in AI model training and development, such as dataset curation before and during the training phase.³ The new TDM exceptions coexist with earlier provisions, notably the exception for temporary acts of reproduction under Article 5(1) of the InfoSoc Directive.⁴ However, their precise relationship with existing AI legislation is complex.⁵ The distinction between these exclusions is unclear.⁶ The temporary reproduction exception, though applicable to transient analysis, fails to address TDM requiring extended data retention.⁷ comprehensively

Regarding TDMs' lawful access criterion, exceptions call for the user to possess 'lawful access' to the material being mined.⁸ However, neither the CDSM Directive nor broader EU copyright law clearly defines lawful access. Though including subscription or open access content, lawful access remains unclear, especially for content accessed via paywalls or illegal sources.⁹ This ambiguity creates uncertainty, especially for commercial AI training where the legality of data acquisition is paramount.¹⁰

Opt-out option questions persist about its scope and timing.¹¹ This system aims to empower rightsholders to control, license, or prohibit TDM of their works.¹² However, the directive implies that technological anti-TDM methods, such as masks or 'poisoned' information, would not constitute genuine rights reservations if they merely deceive TDM agents rather than representing the rightsholder's wishes.¹³ While GenAI providers are held accountable under the AI Act for adhering to the opt-out, implementing it in practice may be challenging

³European Copyright Society, 'Copyright and Generative AI: Opinion of the European Copyright Society' (January 2025) <https://ssrn.com/abstract=000000>, 5.

⁴Eleonora Rosati, *Judge-made EU copyright harmonisation: the case of originality* (PhD thesis, European University Institute 2012), 1.

⁵Mezei P, 'A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI' [2024] SSRN Electronic Journal, 6.

⁶atti, L. "The European solution for text and data mining: A focus on the opt-out system provided by article 4 of the DSM directive," *EUROPEAN INTELLECTUAL PROPERTY REVIEW* (2021), 43(12), pp. 765–770, 6.

⁷Kretschmer M and Margoni T, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' [2021] SSRN Electronic Journal, 10.

⁸R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 *European Intellectual Property Review*, 2021/5, p. 322-337, 12.

⁹vensson G 'Text and Data Mining in EU Copyright Law' (Uppsala University, 2020), 26.

¹⁰Margoni T, 'TDM and Generative AI: Lawful access and opt-outs'[2025] SSRN Electronic Journal, 12.

¹¹Mezei P, 'A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI' [2024] SSRN Electronic Journal, 1.

¹²Gatti, L. "The European solution for text and data mining: A focus on the opt-out system provided by article 4 of the DSM directive," *EUROPEAN INTELLECTUAL PROPERTY REVIEW* (2021), 43(12), pp. 765–770, 5-6.

¹³Mezei P, 'A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI' [2024] SSRN Electronic Journal, 11.

for models trained outside the EU.¹⁴ Consequently, there is doubt over whether the opt-out helps individual writers in a significant way.¹⁵

Technological Protection Measures (TPMs) significantly hinder the use of TDM. The process under Article 6(4) of the InfoSoc Directive allows users to benefit from exceptions despite TPMs being inefficient, complicated, and an extra burden.¹⁶ Users often need to contact rightsholders individually or navigate laborious procedures for each work.¹⁷ The CDSM Directive did not expressly prohibit TPMs from overriding exceptions. A lawful access can arise if contractual terms, especially Article 4 opt-out mechanisms, restrict access despite TDM exceptions. Contracts can thus offer content owners a more effective, adaptable mechanism than IP rights to restrict TDM, particularly for non-Database Directive material. Contracts can provide content owners with a more effective and adaptable mechanism than IP rights to restrict TDM, especially for material not covered by the Database Directive.¹⁸ Uncertainty persists as these contractual measures differ from typical TPMs, and their proportionality is often evaluated against system security rather than the balance with copyright exceptions.¹⁹

Furthermore, it is unclear how the Software Directive and the DSM Directive's TDM exceptions relate to one another in the interaction with the Software Directive.²⁰ The reason that Article 3 does not explicitly include software, whereas Article 4 does, is not apparent. If Article 3 doesn't apply, it raises the question of whether commercial companies can use TDM software only under Article 4, potentially forcing SMEs and start-ups into software licensing.

¹⁴Dusollier S. (PRIA Centre for Regulation of the Creative Economy 2025) working paper, University of Glasgow, 6.

¹⁵European Commission, Directorate-General for Research and Innovation, 'Standardisation in the area of innovation and technological development, notably in the field of text and data mining: report from the Expert Group' (Publications Office 2014) <https://data.europa.eu/doi/10.2777/71122> accessed [2 June 2025], page 28.

¹⁶R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337, 22.

¹⁷Margoni T and Kretschmer M, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685, 32.

¹⁸R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337, 36.

¹⁹R. Ducato, A. Strowel, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, CRIDES Working Paper Series no. 1/2021; forthcoming in 43 European Intellectual Property Review, 2021/5, p. 322-337, 26.

²⁰Maryna Manteghi, 'The insufficiency of the EU's text and data mining exceptions for using artificial intelligence' [2022] 44(11) EUROPEAN INTELLECTUAL PROPERTY REVIEW 651-663, 11.

Finally, all TDM exceptions are subject to the 'Berne three-step test', requiring that exceptions be applied only in certain exceptional cases, not conflict with a normal exploitation of the work, and do not unreasonably prejudice the legitimate interests of the rightholder.²¹ This test's flexible and open-ended requirements introduce significant complexity and legal ambiguity. Rightsholders can strategically use these standards to contest TDM exemption compliance, potentially obstructing national AI initiatives.

Jeroen Zweers' findings on 'low legal data publishing rates' in the Netherlands provide crucial empirical context for these theoretical limits.²² He notes that only 4–5% of Dutch court cases are released, with most being handwritten for anonymisation, making it difficult for projects like GPT-NL to gather sufficient training data.²³ While TDM exceptions aim to facilitate AI training data acquisition, Zweers highlights a critical supply-side constraint: non-copyright factors, such as cost, political will, and institutional resistance, significantly restrict the availability of high-quality, usable data. Thus, despite their legal existence, TDM exceptions are effectively limited by the scarcity of real-world data, posing a serious barrier to AI research in the EU.

III. Economic Perspectives

For decades, the EU has balanced copyright holder incentives with user advantages.²⁴ This means incentivising work production while avoiding limitations that impede helpful societal use, like building GenAI models that rely on freely accessible data.²⁵ Thus, the commercial use of copyrighted content, facilitated by exceptions, is justified if it ensures access to societal information. The EU Commission's TDM Expert Report argues that commercial TDM has high economic value, yielding substantially greater welfare increases than non-commercial TDM.²⁶ It is claimed that, from a financial standpoint, it is meaningless to

²¹Senftleben M, 'Compliance of National TDM Rules with International Copyright Law – an Overrated Nonissue?' [2022] SSRN Electronic Journal, 3.

²² Interview with Jeroen Zweers (Tilburg, 22 May 2025).

²³ibid

²⁴European Commission, Directorate-General for Research and Innovation, 'Standardisation in the area of innovation and technological development, notably in the field of text and data mining: report from the Expert Group' (Publications Office 2014) <https://data.europa.eu/doi/10.2777/71122> accessed [2 June 2025], 27.

²⁵Senftleben M and Buijtelar L, 'Robot Creativity: An Incentive-Based Neighboring Rights Approach' [2020] SSRN Electronic Journal, 13.

²⁶European Commission, Directorate-General for Research and Innovation, 'Standardisation in the area of innovation and technological development, notably in the field of text and data mining: report from the Expert Group' (Publications Office 2014) <https://data.europa.eu/doi/10.2777/71122> accessed [2 June 2025].

discriminate between commercial and non-commercial use because both may improve societal welfare and ought to be encouraged.²⁷

Scholars argue that the content industry influenced the existing system, tightening protections by expanding rights and narrowing exceptions. This could hinder obtaining information and facts from copyrighted works.²⁸ Others view the EU's TDM approach, particularly Article 4's opt-out for commercial uses, as overly restrictive compared to jurisdictions with more liberal provisions. This perceived restriction may render EU Member States less competitive in the high-tech sector. EU businesses could face disadvantages from the opt-out model's higher transaction costs and licensing requirements compared to broader exceptions elsewhere.²⁹

Without a practical economic method for objective revenue benchmarking, while some prefer any compensation to none, economists may argue it compromises societal welfare.³⁰

Although some contend that any compensation is preferable to none at all, economists may counter that this compromises societal welfare.³¹ There is a chance that market failure may become ingrained in legal precedent.³² The use of an input-based strategy runs the risk of harming local high-tech businesses and prompting them to seek better training conditions abroad. This might result in a 'lose-lose' situation if EU trainers relocate overseas and EU rightsholders are not compensated. There is 'conceptual dissonance' between the existing law's emphasis on the input stage and rightsholders' worries about economic exploitation of potentially rival AI outputs.³³

Yang and Zang analyse social welfare policy incentives, examining the relationship between copyright laws, revenue, and overall societal welfare. Their analysis compares liberal versus strict fair use to differentiate short-term and long-term effects.³⁴ The study compares liberal fair use with strict fair use. Under a data-limited regime, the scenario changes: while liberal fair use might be advantageous when training data is plentiful, it could reduce producers'

²⁷ibid

²⁸Margoni T and Kretschmer M, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685, 6.

²⁹European Commission, Directorate-General for Research and Innovation, 'Standardisation in the area of innovation and technological development, notably in the field of text and data mining: report from the Expert Group' (Publications Office 2014) <https://data.europa.eu/doi/10.2777/71122> accessed [2 June 2025], 38.

³⁰Martens B, 'Why the Scope of EU Copyright Protection Should Be Reduced to Realize the Innovation Benefits of Generative AI' [2024] SSRN Electronic Journal, 2.

³¹ibid

³²Senfleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 22.

³³ibid

³⁴Yang SA and Zhang AH, 'Generative AI and Copyright: A Dynamic Perspective' [2024] SSRN Electronic Journal, 4.

motivation to provide human content when data is scarce.³⁵ Reduced data collection costs counter this data scarcity. Consequently, under a data-limited regime, generous fair use may lead to worse model quality and societal welfare compared to stringent fair use. The analysis also considers AI-copyrightability: generous fair use can be advantageous when AI-copyrightability is minimal, but stronger AI-copyrightability can impede AI progress and social benefit. In the long run, rigorous fair use, especially in a competitive environment, could enhance societal welfare more than permissive fair use under data scarcity.³⁶ This is because strict fair use, even with short-term negative impacts, incentivises producers to create more human material, enhancing long-term model development and societal welfare.

The actual reality of low publishing rates of data and the consequent possibility of ‘klassenjustitie’ (class justice) should also be directly connected to these economic concerns.³⁷ Zweers highlights the lack of transparency in the legal system and the costs associated with data anonymisation as contributors to this data shortage.³⁸ Beyond economic waste, existing monetary policy, with limited opt-outs and insufficient open data incentives, creates a ‘lose-lose’ scenario. This not only hinders EU AI development and competitiveness but also has significant societal consequences. This suggests that economic policies are not aligned with fundamental rights and the broader community welfare objectives.

IV. Conclusion

The lack of regulatory clarity and conceptual coherence in the CDSM Directive and its TDM exemption is illustrated in this chapter. Finding that AI outputs are assessed against copyrightability tests not initially designed for machine-trainable algorithms. This necessitates a distinct legal approach, further complicated by the implementation of the opt-out mechanism.

³⁵ibid.(21).

³⁶ibid. (22).

³⁷Interview with Jeroen Zweers (Tilburg, 22 May 2025).

³⁸ibid

CHAPTER V

Conclusion

I. Subquestion Answers

1. Definition of Originality and The Protection Threshold for Partial Reproductions & the Application to GenAI

EU jurisprudence consistently defines originality as an author's 'own intellectual creation,' necessitating that their free and creative choices demonstrably reflect their personality.¹ The major Infopaq decision set a very low bar for the copyright protection of partial reproductions, even for short excerpts as long as they capture the author's original expression.² The standard is anthropocentric, requiring a human author to exercise creative agency and add a personal touch to the work.³ Applying this standard to GenAI creates debates.⁴ AI-assisted works, which require significant human input, may still be protected by copyright.⁵ In contrast, AI-generated works created automatically with minimal human involvement may not be protected by copyright under current EU law. Meaning reproduction might not be protected, even if it's objectively unique and innovative, unless its expressive features clearly show human creative contribution.

2. Latent knowledge extraction and Partial Copying Requirement according to Infopaq

'Latent knowledge extraction' refers to the method by which AI models examine and absorb patterns, styles, and other fundamental information from large training datasets, as opposed to the explicit archiving or replication of particular copyrighted works.⁶ In the training phase, copyrighted works are statistically transformed into numerical parameters within the model's vector space through a process of digestion.⁷ Although AI models generally do not retain exact replicas of their training data, they are demonstrably capable of 'memorising' and

¹S van Gompel and M van Eechoud, 'Creativity, Autonomy and Personal Touch: A Critical Appraisal of the CJEU's Originality Test for Copyright' in M van Eechoud (ed), *The Work of Authorship* (Amsterdam University Press 2014), 1.

²PB Hugenholtz and JP Quintais, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 *IIC* 1190, 9.

³Bonadio E and McDonagh L, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Sweet and Maxwell 2020), 3.

⁴Fenwick, Mark and Jurcys, Paul, 'Originality and the Future of Copyright in an Age of Generative AI' (February 11, 2023), 7.

⁵Christophe Geiger, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative AI' [2024] 55(7) *IIC* - International Review of Intellectual Property and Competition Law 1129-1165, 10.

⁶Nathan Pelgrims, 'Lockean theories of intellectual property and the product of artificial intelligence' (Master's thesis, Universiteit Gent 2021), 77.

⁷TimW Dornis, 'Generative AI, Reproductions Inside the Model, and the Making Available to the Public' [2025] 56(6) *IIC* - International Review of Intellectual Property and Competition Law 1-30,4.

subsequently 'regurgitating' near-identical or identical portions thereof in their generated outputs.⁸

Instances of such memorisation render AI-generated outputs susceptible to classification as partial copies from a legal standpoint, thereby potentially infringing upon existing protected works.⁹ Meaning, the Infopaq originality requirement demands a qualitative evaluation, focusing on whether a 'substantial part' of the author's creative work has been reproduced.¹⁰ If an output reproduces a portion of an original work that is deemed original due to its 'choice, sequence, and combination of words' or other expressive elements, it may constitute an infringing partial reproduction.¹¹ Even if the AI's internal mechanism isn't precisely equivalent to human copying, objective proof of replication in the AI output is necessary to prove infringement, particularly for complex works where accidental creation is statistically unlikely.¹² Nevertheless, a significant impediment persists for rightholders in adducing evidence of direct derivation from a specific source embedded within a voluminous and opaque training dataset.¹³

3. Effectiveness of EU Copyright Laws in Regulating AI Reproduction & Doctrinal Gaps

The current EU copyright laws' copyright mechanism is separated into two limbs: the input and the output. The first regulates the TDM processes and access to data, while the latter governs the output format. The AI Act further reinforces this by requiring AI providers to establish clear rules respecting these opt-outs and to offer transparent summaries of their training data.¹⁴ While ostensibly preserving traditional copyright values, TDMs requirements risk consigning a vast quantity of AI-generated content to the public domain, thereby potentially disincentivising investment in autonomous AI research.¹⁵ Secondly, the efficacy of

⁸Cooper, A. Feder and Grimmelmann, James, The Files are in the Computer: Copyright, Memorization, and Generative AI (April 22, 2024). Cornell Legal Studies Research Paper Forthcoming, Chicago-Kent Law Review, Forthcoming, 2.

⁹Quintais JP, 'Generative AI, Copyright and the AI Act' [2025] 56 - Computer Law & Security Review 106107, 4.

¹⁰Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, 58.

¹¹Synodinou, Tatiana, Book Chapter: The Foundations of the Concept of Work in European Copyright Law (January 30, 2012). Synodinou (ed.), "Codification of European Copyright, Challenges and perspectives, the Hague, Kluwer Law International, p. 93-113, 2012, 8.

¹²Guadamuz A, 'A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs' [2023] SSRN Electronic Journal, 19.

¹³Christophe Geiger and Vincenzo Iaia, 'The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI' [2024] 52(1) Computer Law & Security Review: The International Journal of Technology Law and Practice, 10.

¹⁴Regulation (EU) 2024/1689 of the European Parliament and of the Council of 10 May 2024 on a European framework for AI, OJ L 1689, 10.5.2024, p. 1, art. 53(1)(c).

¹⁵Fenwick, Mark and Jurcys, Paul, Originality and the Future of Copyright in an Age of Generative AI (February 11, 2023), 30.

the TDM exceptions is debated by scholars.¹⁶ First, Article 4 opt-out aims to empower rightholders, but its effectiveness is undermined by the immense scale of training data and the absence of standardised, machine-readable protocols, resulting in prohibitive 'transaction costs' for licensing. Creating a difficulty for AI developers, often without ensuring fair compensation for individual creators.¹⁷ Economically speaking, this approach is viewed as overly protective, which could impede AI innovation in the EU by raising costs and dividing data access, thereby putting EU AI companies at a competitive disadvantage in comparison to US and Japanese regulations.¹⁸ Thirdly, there are still major issues with enforcement and transparency.¹⁹ Notwithstanding the AI Act's transparency requirements, which are frequently criticised for their insufficient specificity, establishing direct infringement stemming from 'latent knowledge extraction' or AI output regurgitations remains arduous.²⁰ This inherent opacity perpetuates a 'black-box' issue, wherein the intricate internal workings of AI algorithms obscure potential copyright infringement. This fundamental conflict highlights that the EU's copyright system, initially designed for traditional copying and human creation, is demonstrably ill-equipped to address the evolving creative and economic realities of generative AI effectively. Consequently, this mandates comprehensive legislative reform to explore alternative regulatory models, including statutory licensing or levies on AI output, with the objectives of ensuring equitable compensation for creators, fostering AI innovation, and mitigating the existing difficulty of proving infringement.

II. Main Question Answer

1. The Approach to Originality and The Gaps in the EU Framework

The CJEU's well-established low originality standard in partial copying, rooted in human intellectual creation, presents significant drawbacks when applied to the intricate processes of generative AI's output production and latent knowledge extraction. As emerging AI models are capable of mimicking human behaviour to the extent of their abilities based on detectors, it raises the question of whether the standard, established long ago, needs amendment to

¹⁶Maryna Manteghi, 'The insufficiency of the EU's text and data mining exceptions for using artificial intelligence' [2022] 44(11) EUROPEAN INTELLECTUAL PROPERTY REVIEW 651-663, 12.

¹⁷Christophe Geiger, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative AI' [2024] 55(7) IIC - International Review of Intellectual Property and Competition Law 1129-1165, 23.

¹⁸Senftleben M, 'Ai Act and Author Remuneration - A Model for Other Regions?' [2024] SSRN Electronic Journal, 12.

¹⁹Dusollier S. (PRIA Centre for Regulation of the Creative Economy 2025) working paper, University of Glasgow, 11.

²⁰Buick A, 'Copyright and AI Training Data—Transparency to the Rescue?' (2024) 20 Journal of Intellectual Property Law and Practice 182, 1.

protect authors, users, and research purposes beyond commercial incentives, as the EU strives to promote. A conflict in economic policy further exacerbates this issue, particularly the commercial opt-out, which is perceived by some as leading to overprotection that stifles AI innovation and creates inefficient data markets, often without ensuring equitable compensation for individual creators. Although the existing framework successfully encourages ‘human-in-the-loop’ creativity, it struggles to manage the financial implications of AI's data-intensive training and the increasing autonomy of its outputs. The ultimate goal is to establish a new balance that supports both technological advancement and human ingenuity simultaneously. Therefore, by striving to bridge the gap between human and AI knowledge extraction techniques, it opens another one. A licensing and economic gap puts the EU a step behind its competitors in a field that is growing faster by the day.

The proposed amendments to these training, originality and economic gaps deriving from the EU's Copyright Law include:

- a) More flexible DSMD provision for TDM, given the shortage of data available. By restricting a developing good, by a second variable, the risk of progressiveness and the digital future of the EU becomes limited. First, due to the open data available, along with the further restriction of subsequent access-based systems. Second, due to the extraterritoriality of training, excluding the field of use in the EU, if the training or subsequent updates do not adhere to the TDM exceptions of the EU, by chance. The Japanese model strikes a balance between these two variables while preserving author integrity and promoting progress. Taking inspiration from the EU's perspectives would optimally open the door more for companies, developers, and users.
- b) Clearer international competitive licensing structures with an author-based registry to solve two problems. First, mitigate the risk of an investor-based monopoly by establishing a set price range system, thereby clearing the field for big developing companies. Second, provide a system of reach and reward for authors, eliminating the problems of contacting and consequently claiming non-responses by developers. This would create a clear market-based system, facilitating the values of copyright as an IP monopoly while preserving the competitive and integral values of the EU.

III. Final Remarks and Limitations of The Research

This study has the following limitations restricting the conclusiveness of its findings:

1. The study's research is impeding the originality standards, limiting the input and output generated content, while not examining any other data or privacy issues affecting the access, processing or production of the data.
2. The analysis did not conduct a comparative legal analysis with other major jurisdictions, such as the US or Japan.
3. There is a lack of thorough empirical data on the real economic effects of TDM exceptions and AI-generated outputs, such as direct effects on creator compensation, market displacement, or investment levels, which reduces the capacity to make firm policy judgements based only on empirical data.

The questions posed by this study will be effectively answered in the judgment of the ongoing case of *Like Company v Google Ireland*.²¹ The questions posed in the proceeding will effectively answer the requirements for constituting a reproduction, the scope of the TDM exception, and the responsibility of developers for copyrighted production.

²¹Court of Justice of the European Union, 'Summary of the request for a preliminary ruling in Case C-250/25' (3 April 2025) <https://curia.europa.eu/juris/showPdf.jsf?docid=300681&doclang=EN>.

Bibliography

Cases

1. *Bezpečnostní softwarová asociace - Svaz softwarové ochrany v Ministerstvo kultury* C-393/09 ECLI:EU:C:2010:816
2. *BSA v Ministry of Defence* EWHC Ch 1
3. *Football Dataco Ltd v Sportradar GmbH* C-604/10 ECLI:EU:C:2012:300
4. *Football Dataco Ltd v Yahoo! UK Ltd* EWHC 1400 (Ch)
5. *Frank Peterson v Google LLC and YouTube LLC* C-682/18 ECLI:EU:C:2021:503
6. *Infopaq International A/S v Danske Dagblades Forening* C-5/08 ECLI:EU:C:2009:465
7. *Painer v Standard VerlagsGmbH and Others* C-145/10, Opinion of AG Trstenjak, ECLI:EU:C:2011:798
8. *Pelham GmbH v Ralf Hütter* C-476/17 ECLI:EU:C:2019:624
9. *Pelham GmbH and Others v Ralf Hütter and Florian Schneider-Esleben* C-476/17 EU:C:2019:624
10. *The Authors Guild and Others v OpenAI Inc. and Others*, Case No 1:23-cv-08292 (SDNY, filed 19 September 2023)
11. *The New York Times Company v Microsoft Corporation and Others*, Case No 1:23-cv-11195 (SDNY, filed 27 December 2023)

Legislation

1. Berne Convention for the Protection of Literary and Artistic Works (signed 9 September 1886, entered into force 5 December 1887) 828 UNTS 221
2. 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 OJ L167/10
3. Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure OJ L157/1
4. 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 OJ L130/92
5. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 10 May 2024 on a European framework for AI OJ L1689/1

Books and Book Chapters

1. Bonadio E and McDonagh L, *Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity* (Sweet and Maxwell 2020)
2. Gompel S van and Eechoud M van, 'Creativity, Autonomy and Personal Touch: A Critical Appraisal of the CJEU's Originality Test for Copyright' in Eechoud M van (ed), *The Work of Authorship* (Amsterdam University Press 2014)
3. Synodinou T, 'The Foundations of the Concept of Work in European Copyright Law' in Synodinou T (ed), *Codification of European Copyright, Challenges and Perspectives* (Kluwer Law International 2012)

Journal Articles

1. Begishev IR, Latypova EY and Kirpichnikov DV, 'Artificial Intelligence as a Legal Category: Doctrinal Approach to Formulating a Definition' (2020) 14 *Actual Problems of Economics and Law* 79
2. Buick A, 'Copyright and AI Training Data—Transparency to the Rescue?' (2024) 20 *Journal of Intellectual Property Law and Practice* 182
3. Buick A, 'Copyright and AI Training Data—Transparency to the Rescue?' (2025) 20(3) *Journal of Intellectual Property Law and Practice* 182
4. Burk DL, 'Thirty-Six Views of Copyright Authorship, by Jackson Pollock' (2020) *Journal of World Intellectual Property* 320
5. Carroll MW, 'Copyright and the Progress of Science: Why Text and Data Mining Is Lawful' (2019) 53 *University of California, Davis, Law Review* 893
6. Ceci F, Masciarelli F and Poledrini S, 'How social capital affects innovation in a cultural network: Exploring the role of bonding and bridging social capital' (2019) 17 *European Journal of Innovation Management* 973
7. Ferri F, 'The Dark Side(s) of the EU Directive on Copyright and Related Rights in the Digital Single Market' (2020) 7 *China-EU Law Journal* 21
8. Geiger C and Iaia V, 'The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI' 52(1) *Computer Law and Security Review* 7
9. Geiger C, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative

- AI' 55(7) IIC - International Review of Intellectual Property and Competition Law 1129
10. Geiger C, 'The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the Failed Text and Data Mining Limitations in the CSDM Directive' SSRN Electronic Journal
 11. Hugenholtz PB and Quintais JP, 'Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?' (2021) 52 IIC 1190
 12. Kroeze LJ, 'Legal Research Methodology and the Dream of Interdisciplinarity' (2013) 16(3) Potchefstroom Electronic Law Journal 13
 13. Lemley M, 'How Generative AI Turns Copyright Law Upside Down' (2024) 25 Science and Technology Law Review 6
 14. Manteghi M, 'In Search of Balance: Text, Data Mining and Copyright in the Digital Single Market Directive from a Fundamental Rights Perspective' (2023) 48(4) European Law Review 443
 15. Maryna Manteghi, 'The insufficiency of the EU's text and data mining exceptions for using artificial intelligence' (2022) 44(11) European Intellectual Property Review 651
 16. Bottis M, Papadopoulos M, Zampakolas C and Ganatsiou P, 'Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit' (2019) 12 Erasmus Law Review 190
 17. Nordemann JB and Pukas J, 'Copyright Exceptions for AI Training Data—Will There Be an International Level Playing Field?' (2022) 17(12) Journal of Intellectual Property Law and Practice 973
 18. Quintais JP, 'Generative AI, Copyright and the AI Act' 56 Computer Law and Security Review 106
 19. Rosati E, 'Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity' (2019) 27 Asia Pacific Law Review 198
 20. Rosati E 'Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law' European Journal of Risk Regulation 1

Online Sources and Reports

1. AIPPI German Delegation, 'Copyright in Artificially Generated Works' (National Report, 2019) <https://aippi.soutron.net/Portal/Default/en-GB/RecordView/Index/254>
2. Bomhard D and Siglmüller J 'AI Act—das Trilog Ergebnis' (2024) Recht Digital 45

3. Cooper A, Feder A and Grimmelmann J, 'The Files are in the Computer: Copyright, Memorization, and Generative AI' (Cornell Legal Studies Research Paper, 22 April 2024); forthcoming in Chicago-Kent Law Review
4. Court of Justice of the European Union, 'Summary of the request for a preliminary ruling in Case C-250/25' (3 April 2025)
<https://curia.europa.eu/juris/showPdf.jsf?docid=300681&doclang=EN>
5. Drexel J and others, 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Perspective' SSRN Electronic Journal
6. Ducato R and Strowel A, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out' (CRIDES Working Paper Series no 1/2021, 2021); forthcoming in (2021) 43 European Intellectual Property Review 322
7. Dusollier S, 'Untitled working paper' (PRIA Centre for Regulation of the Creative Economy, University of Glasgow 2025)
8. European Commission, *Shaping Europe's Digital Future* COM(2020) 67 final (19 February 2020)
9. European Commission, Directorate-General for Research and Innovation, 'Standardisation in the area of innovation and technological development, notably in the field of text and data mining: report from the Expert Group' (Publications Office 2014)
<https://data.europa.eu/doi/10.2777/71122>
10. European Copyright Society, 'Copyright and Generative AI: Opinion of the European Copyright Society' (January 2025)
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5166831
11. Fenwick M and Jurcys P, 'Originality and the Future of Copyright in an Age of Generative AI' (11 February 2023)
12. Feuerriegel S and others, 'Generative AI' SSRN Electronic Journal
13. Flynn S and others, 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action' SSRN Electronic Journal
14. Geiger C, Frosio G and Bulayenko O, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' SSRN Electronic Journal
15. Gwagwa A, (The Royal Institute of International Affairs Chatham House 2024) research paper
16. Guadamuz A, 'A Scanner Darkly: Copyright Infringement in Artificial Intelligence Inputs and Outputs' SSRN Electronic Journal
17. Kretschmer M and Margoni T, 'A Deeper Look into the EU Text and Data Mining

Exceptions: Harmonisation, Data Ownership, and the Future of Technology' SSRN Electronic Journal

18. Margoni T and Kretschmer M, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71 GRUR International 685
19. Margoni T, 'Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?' SSRN Electronic Journal
20. Margoni T, 'TDM and Generative AI: Lawful access and opt-outs' SSRN Electronic Journal ²
21. Martens B, 'Why the Scope of EU Copyright Protection Should Be Reduced to Realize the Innovation Benefits of Generative AI' SSRN Electronic Journal ²
22. Maslej N and others, *The AI Index 2025 Annual Report* (Stanford University, Institute for Human-Centered AI, April 2025) <https://ai100.stanford.edu/>
23. McDaniels Law, 'First Case On Whether AI Fits Into EU Copyright Law Referred To The CJEU' (McDaniels Law, 2024) <https://mcdanielslaw.com/first-case-on-whether-ai-fits-into-eu-copyright-law-referred-to-the-cjeu/>
24. Mezei P, 'A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI' SSRN Electronic Journal
25. Peukert, 'Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training' (Research Paper of the Faculty of Law of Goethe University Frankfurt/M. No 3/2024, 2024)
26. William Fry, 'CJEU to Rule on AI Chatbots and Copyright in Landmark Case Against Google' (William Fry, 2024) <https://www.williamfry.com/knowledge/cjeu-to-rule-on-ai-chatbots-and-copyright-in-landmark-case-against-google/>
27. Yang SA and Zhang AH, 'Generative AI and Copyright: A Dynamic Perspective' SSRN Electronic Journal

Theses and Dissertations

1. Barnwell M, 'Balancing the benefits of TDM against copyright protection: A

comparative study on the exceptions and limitations to the reproduction right with regard to text and data mining in the EU and the US' (LLM Thesis, Tilburg University 2018)

2. Pelgrims N, 'Lockean theories of intellectual property and the product of artificial intelligence' (Master's thesis, Universiteit Gent 2021)
3. Rosati E, 'Judge-made EU copyright harmonisation: the case of originality' (PhD thesis, European University Institute 2012)
4. Svensson G, 'Text and Data Mining in EU Copyright Law' (Uppsala University, 2020)