

GENERATIVE AI AND EU COPYRIGHT LAW:

Exploring Exceptions and the Derivative Works Concept

Clemens Danda

AUTHOR

The author is a postdoctoral researcher focusing on Private Law, Law of Technology, and Cultural Property Law. He holds a PhD in Private Law from the University of Vienna, an LL.M. in IP & ICT Law from KU Leuven in Brussels, and a MA in Art History from the Technical University Berlin.

TABLE OF CONTENTS

I. INTRODUCTION	35
II. THE DATA MINING EXCEPTION	35
A. A Brief Introduction to the Exception	35
B. Applying the TDM Exception to Generative AI	36
C. Challenges of the Data Supply Chain	38
D. Interim Conclusion	38
III. THE EXCEPTION FOR TEMPORARY OR INCIDENTAL REPRODUCTION	39
A. The Requirements of the Exception	39
B. The Subject Matter Covered by the Exception	39
IV. NORMAL USE AND EXCEPTIONS UNDER THE DATABASE DIRECTIVE	39
V. OUTPUT AS LAWFUL DERIVATIVE WORK	40
A. Copyright Exceptions and Commercial Deployment	40
B. Harmonizing EU Law and National Concepts	42
C. Distance to Protected Works	42
D. Originality of the Derivative Work	44
VI. CONCLUSION	46

I. INTRODUCTION

Generative AI poses numerous challenges to EU copyright law. In this inquiry I will mainly focus on two critical issues: The usage of copyrighted materials for the development of generative AI models and the subsequent act of generating digital content (images, video, audio, or text). The development of generative AI models mainly involves data scraping, data preparation, as well as the training of the model. In regard of these acts, we will assess the applicability of existing copyright exceptions for text and data mining, temporary or incidental reproduction, as well as exceptions related to database rights. When it comes to the publication of generated digital content, our focus will be on determining the conditions under which such content can be recognized as a legally permissible derivative work in relation to the copyrighted originals.

II. THE DATA MINING EXCEPTION

A. A Brief Introduction to the Exception

The goal of the first text and data mining (TDM) exception in Art. 3 of the Directive on Copyright in the Digital Single Market (CDSMD)¹ is to enable research organizations and cultural heritage institutions to reproduce, extract, and analyze text and data provided they have already obtained some form of access license (e.g. academic databases) or when such data is freely accessible online (Art. 3 (1), Rec. 11, 14 and 18). The exception aligns with a long tradition of research exceptions in other contexts acknowledged internationally.² Additionally, under the condition that the use for TDM has not been reserved by the rightsholder (opt-out), Art. 4 permits private entities and other institutions to TDM. Both exceptions generally provide the needed clarification on the question whether TDM is permissible as a new form of use respectively under which conditions it can be conducted (Rec. 8). An important point is that the Directive differentiates, on the one hand, between works protected under copyright or data base rights and, on the other hand, unprotected ideas, facts or data. In doing so, it follows the fundamental principle of copyright law that ideas, facts and data are not subject to protection. The Directive therefore does not introduce data property (Rec. 9 “in such instances no authorization is required under copyright law”).

Despite the obvious advantage of having clarification in this regard, the exceptions has been criticized from the beginning. Some argue that there is no actual need for a TDM exception since the underlying processes of TDM as well as training AI models supposedly do not resemble any use protected under copyright. In their view, there is no such thing as a mining or training right in need of restriction. For example, the European Copyright Society states that “TDM has no impact on the normal exploitation of works or other protected content. [...] Works which are subject to TDM are not used as works”³. Specific to AI model training, others added that residues of text or data contained within the model –

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.

² For a comparative overview of research exceptions see Sean Flynn et alia, *Research Exceptions in Comparative Copyright*, PIJIP/TLS Research Paper Series no. 75 (2022), available from <https://digitalcommons.wcl.american.edu/research/75>.

³ European Copyright Society, *General Opinion on the EU Copyright Reform Package 5* (European Copyright Society 2017); compare Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, *Journal of the Copyright Society of the USA*, Vol. 66 (2019), 291 who argues for US law that “copying expressive works for non-expressive purposes should not be counted as infringement and must be recognized as fair use.”

such as in tokens or weights – constitute no reproductions of the original works but merely represent reproductions of unprotected data contained within protected works. In their view, such extracted data can neither fulfill the requirement of originality nor the idea/expression dichotomy.⁴ The criticism stems in part from the already mentioned opt-out rule established in Art. 4(3) which enables rightsholders to fully restrict TDM. Margoni/Kretschmer comment in this regard, that closing off parts of the available data sources ultimately contradicts the EU's goals on digital innovation, since it disincentivizes stakeholders to develop AI systems within the EU.⁵ The criticism has some weight to it, since human made text and data are necessary for the development of most models. Alternatives such as synthetic data potentially come with performance downsides.⁶ Keeping these points in mind, we will try to apply the TDM exceptions onto generative AI. The critical question at hand is whether the exceptions adequately adapt to the latest technological advancements.

B. Applying the TDM Exception to Generative AI

The scope of the TDM exceptions enable reproduction, extraction from databases, data normalization, as well as the retainment of copies as long as necessary for TDM (Rec. 8-10, Arts. 3(2), 4(2) CDSMD). They do not address, however, further details of the TDM (or the AI development) process, implementing a technology neutral approach. In particular, they do not provide which acts related to data preparation are covered except for the mentioned *normalization*. Model development requires extensive preparation of data by processes such as normalization, conversion, standardization. Finally, the training of the AI model based on the adapted text or data is not mentioned by the CDSMD.⁷ How should we handle the gap between the TDM practices known in 2018 and the recent use case of generative AI?

In regard of the development of AI models, we should first consider that Art. 2(2) CDSMD defines TDM very broadly as “any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations”. The phrasing “any automated analytical technique” certainly points at a high flexibility regarding new technical applications.⁸ We could argue that, since AI models are analytical tools, the CDSMD should be interpreted as allowing training as well as the generation of data output. However, the exceptions were mainly conceptualized for research or internal knowledge management purposes. Recital 8 speaks of „gaining new knowledge and discovering new trends“. However, as cited above, TDM also serves to “generate information which includes but is not limited to patterns, trends and correlations” (Art 2(2) CDSMD). While this definition is broad, it remains challenging to determine whether the exception applies to the generation of images, sounds, or audiovisual works intended for entertainment or commercial marketing purposes. In my point of view, the inclusion of such output can be supported based on the following points: As mentioned before, Art. 2(2)

⁴ Thomas Margoni & Martin Kretschmer, A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, 700 GRUR International, 71, 8 (2022).

⁵ Ibid 8.

⁶ Iliia Shumailov et alia, The Curse of Recursion: Training on Generated Data Makes Models Forget, arXiv:2305.17493 (2023).

⁷ Compare Jan-Torben Evers & Allesandro Buttignon, MLOps (Machine Learning Operations) – Der lange Pfad vom PoC zum Deployment, 19 ITSpektrum (2023) 3.

⁸ Cf. Thomas Margoni & Martin Kretschmer, A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, 687 GRUR International, 71, 8 (2022).

emphasizes that the scope of information that can be generated is not limited to specific categories. Additionally, recital 18 additionally provides that the TDM exception can be used to analyze data “in different areas of life and for various purposes [including] new applications or technologies”. Lastly, taking the general openness of the Directive towards innovation into account (Rec. 2 and 5 CDSMD),

the TDM exceptions appear as dynamic instruments which can address all kinds of data analysis procedures including AI development purposes. This should even apply when the development of generative AI is geared towards entertainment products or marketing objectives. Taking these arguments into account, it is reasonable to apply the exceptions to generative AI as long as their usage remains restricted to internal purposes by the private entity or institution. An extension to publication of output (beyond scientific results) is instead not covered by the Directive and cannot be inferred based on the arguments above.

Consequently, since the development of models generally fall under the TDM process, the argument extends to reproduced data residues within a model (e.g. tokens, weights). Beyond such teleological reasoning, the notion that statistical data are not protected by copyright following the idea/expression dichotomy can be supported.⁹ Rejecting protectability of statistical data appears justified in light of the CJEU’s ruling on the protection of parts of works in *Infopaq*¹⁰. For a part to be eligible for protection, the court either requires that the part shares the overall originality of the protected work or represents an original work in itself.¹¹ The residues in the model do neither represent the whole work (or its originality) nor a relevant part. The training process, instead, establishes a complex statistical relationship model among non-protectable individual words or visual and acoustic data points, which themselves are not integrated into the model. I think, we tend to be conceptually misled by the black box phenomenon here. We cannot easily conceive that the model is able to provide output that closely resembles the original input without retaining a copy in any traditional or even digital sense. The process certainly challenges (or breaks with) our conventional understanding of reproduction. Lastly, depending on the individual expression, the generated output may by itself infringe upon the copyright of a protected work, whether it's the input data or any other protected works. However, given that generating output is an essential component of Text and Data Mining (TDM), we should assert that it does not constitute an infringement as long as it is exclusively used for internal purposes. The question of the output and its publication will be examined in more detail in section V on derivative works.

The Directive provides only a vague case-by-case rule on the permissible retention period of text and data reproductions. Under Art. 4(2) it can be stored “as long as necessary for the purposes of text and data mining”. Since the model itself, the data residues within the model, as well as the generated output can be understood as derivatives to the original data and its reproductions, the rule could potentially apply to all of them. However, as argued above, the model as well as data residues merely resemble statistical relations between words and other data and therefore do not reproduce the protected original. Moreover, the rule should only cover such output that closely reproduces an original work and does not count as a permissible derivative (see V.). Regarding the period of necessity, the manifold types of TDM leave it unclear at which point in time the reproductions need

⁹ Thomas Margoni & Martin Kretschmer, A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology, 690 *GRUR International*, 71, 8 (2022).

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

¹¹ *Ibid.*, [38, 39].

to be deleted. Prolonged retention time for the purpose of explainability of AI or other compliance requirements must certainly be considered. Consistency with the requirements of the GDPR, especially with regard to storage limitation as well as purpose limitation and secondary use, is advisable when personal data are also involved. However, addressing these aspects in detail would require further examination of the GDPR, as well as the unfinished AI Act¹² and the Data Act¹³ and cannot therefore not be addressed in detail here.

C. Challenges of the Data Supply Chain

Data crawling (reproduction) and data preparation (adaptation) appear as more problematic acts than the actual training of the model (analysis). Even more problematic is that data crawling and data preparation are often outsourced to third parties. Professional data collectors will grant access to or transfer data sets to developers. The Directive shortly touches upon such third-party involvement in the context of the research exception (Rec. 11) but does not address outsourcing of data collection or acquisition. It is likely that the entire supply chain, including various forms of data providers, will need to adhere to the requirements of the TDM exceptions. This follows from the CJEU's reasoning in the B2C cases *Filmspelers*¹⁴ and *GS Media*¹⁵ which concern access to unlicensed material within the context of Article 5(1)(b) ISD. In these scenarios, establishing a compliant supply chain is imperative when acquiring copyrighted works to fulfill the "permission by law" criterion. This logic could apply to the realm of data scrapping by professional data providers. It would prevent any TDM institutions – which themselves fulfill the requirements of the exceptions and have a license in regard of the received data (!) – from acquiring data sets from a non-compliant data provider. The issue becomes even more difficult considering the above-mentioned rule for retention of data “as long as necessary for the purposes of text and data mining”. It seems questionable to assume that the necessary retention time is unlimited for independent data providers. Amendments and clarifications by the EU legislator are greatly needed here.

D. Interim Conclusion

So far, we can note that the CDSMD allows for the development of AI models including generative AI. Moreover, based on a teleological interpretation, those institutions and private entities which create the model can generate output including text, image, sound, or audiovisual, output for their own use including development of consumer entertainment products. They may even retain generated output that contains protected elements of the original as long as storage is necessary for the development. Unfortunately, the exception does not sufficiently address the complexities of data supply chain, which involves multiple stakeholders responsible for data collection, preparation, and training tasks. Finally, the Directive does not regulate providing output to third parties – which we will explore later from the perspective of derivative works (below at V.).

¹² Cf. Art. 10(5) AI Act Compromise Draft regarding negative bias detection and correction in relation to the high-risk AI systems, 20(1) and 29(5) for log data retention, and Art. 54 for data retention in AI sandboxes.

¹³ Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/0068 final.

¹⁴ CJEU Case C-527/15 *Stichting Brein v Jack Frederik Wullems* [2017] ECLI:EU:C:2017:300 [69].

¹⁵ CJEU Case C-160/15 *GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker* [2016] ECLI:EU:C:2016:644 [54].

III. THE EXCEPTION FOR TEMPORARY OR INCIDENTAL REPRODUCTION

A. The Requirements of the Exception

Recital 18 CDSMD clarifies that the TDM exception does not prevent public or private entities from relying on other exceptions such as the one for temporary or transient reproduction in Article 5(1) Copyright and Information Society Directive (ISD). Nevertheless, Article 5(1) contains a long list of requirements which challenges its application to AI model development.¹⁶ First, data collection and preparation for AI training is neither transient in the sense of storage for the completion of a distinct technological process¹⁷ nor is its subsequent deletion commonly independent from human intervention.¹⁸ Secondly, the reproduction is not incidental given that copying text and data is the goal of the data collection process. Thirdly, the exception only applies if the data is not modified, a condition that does not align with the data preparation process. Fourthly, it is debatable whether the reproduction has a separate economic value. Ultimately, the value lies in the completed data set, which enables development, thereby creating additional economic utility. Ultimately, in terms of the legality of the data source, counting on implied consent for the purpose of AI development based on deliberate uploading is certainly insufficient – the GDPR will require more specific consent –, and permanent copying does not align with legitimate usage patterns, such as browsing by prospective customers (Cf. Rec 33 ISD).

B. The Subject Matter Covered by the Exception

Besides these hard requirements not being met, the exception was had never been intended for the development of digital products. Recital 33 ISD refers to reproduction of works for efficient transmission in a network, as well as browsing and caching.¹⁹ Hence, the exception is not indifferent towards technological innovation, but instead mostly relates to a different subject matter. Due to the transient nature of text and data usage in the training process, there is a likelihood that the exception could be tailored to suit the training process as a form of lawful use. Nonetheless, isolating the training process from the preceding data collection and preparation process not only appears artificial but also risks ignoring the requirement of a compliant data supply chain.

IV. NORMAL USE AND EXCEPTIONS UNDER THE DATABASE DIRECTIVE

The general idea behind the database copyright right and the *sui generis* data base right of the Database Directive²⁰ is to prevent extractions and (re-)utilizations of either creative databases or substantive investments in databases. Concerning creative databases protected under copyright, various forms of usage of the entire database or substantial parts are protected including temporary reproduction and adaptation (Art. 5). Arguing for AI development based on insubstantiality (Art. 7(1))

¹⁶ Concerning TDM compare DG for Internal Policies, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (2018) 9.

¹⁷ CJEU Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECLI:EU:C:2009:465 [62].

¹⁸ Ibid [64].

¹⁹ Cf. Stavroula Karapapa, Defences to Copyright Infringement: Creativity, Innovation and Freedom on the Internet 116 (2020).

²⁰ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases is a directive of the European Union in the field of copyright law, made under the internal market provisions of the Treaty of Rome.

where the AI developer has a license, seems misplaced given the typically large extent of data necessary for AI development. Instead it appears tempting to interpret *normal use* (Art. 6(1)) as allowing acts of reproduction, adaptation, and communication for the purpose of AI development. However, the general purpose of the exception rather infers an interpretation where *normal use* is confined to use of a database to retrieve data for common informational use or other express terms negotiated by the parties.²¹ Finally, considering that the exceptions listed under Art. 6(2) and Art. 9 – private purpose, scientific research, public concerns, or national law – are exhaustively defined, it is probably not legitimate to dynamically add new forms of usage. *De lege ferenda*, any additional exception would need to align with the requirement that it does not unreasonably prejudice the rightsholder’s legitimate interests or conflicts with normal exploitation of the database (as referred to in Rec. 35 and Art. 6(3)). Licensing of databases for AI development certainly represents a separate economic interest and should not be implied in every access license.

V. OUTPUT AS LAWFUL DERIVATIVE WORK

A. Copyright Exceptions and Commercial Deployment

So far, we looked at the development of AI models and the reproduction of text and data intended for internal usage by the developer. The exceptions discussed above, however, do not encompass the reproduction, communication, or making available of output to third parties (commercial deployment phase). Therefore, generating output that closely resembles training data or other copyrighted works carries the risk of copyright infringement when shared with third parties. Publication of scientific research likely poses an exceptional case in this context. Art. 5(3)(a) of the Information Society Directive (ISD)²² allows EU member states to provide a scientific research exception concerning reproduction and – notably – the right of communication and making available to the public. The prevailing view suggests that the publication of research results falls within this exception.²³ The separate debate over whether allowed acts are confined to mere "illustrations" under Article 5(3)(a)²⁴ is of minor concern in this context, as long as the generative AI does not reproduce substantial parts of the original work. Hence, the central challenge revolves around the reproduction and publication of generated works for the benefit of both commercial enterprises and private individuals which will need to rely on other legal grounds.

Clearly, there is clearly no specific exception available for reproductions or publication made for commercial purposes. In contrast, the reproduction of works for private purposes acquired from lawful

²¹ With the same result DG for Internal Policies, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects (2018) 11.

²² See also Art. 10(1) Rental and Lending Directive and Arts. 6(2)(b), 9(b) Database Directive the latter exception being narrower in scope only referencing extraction but not re-utilization, reproduction or publication, cf. Triaille et alia, Study on the Application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society (2013), 359 et seq.

²³ Triaille et alia, Study on the Application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society (2013), 363.

²⁴ Compare Christina Angelopoulos, Study on Copyright and Scientific Publications: Encouraging Access and Re-use, available from <https://copyrightblog.kluweriplaw.com/2022/12/08/study-on-copyright-and-scientific-publications-encouraging-access-and-re-use/>; Estelle Derclaye, The legal protection of databases, (2008), 137.

sources²⁵, is already harmonized for copyright (Art. 5(2)(b) ISD) as well as for database rights (Art. 6(2)(a), 9(a) Database Directive). In regard of the CJEU's test established in *GS Media*²⁶ which requires that the consumer "did not know or could not reasonably have known" of the illegitimacy of the source of a protected work, the private user of generative AI generally cannot know whether a generative AI model was trained solely on open-source or licensed works. The situation is, however, different when the user can, for example, recognize the residues of metadata, watermarks, or other content identifiers.²⁷ The challenges are different for the provider. In *VCAST Limited*²⁸ the CJEU ruled that a commercial undertaking cannot offer recording services which enable private individuals to remotely create private copies of TV transmissions. The central issue in this case shares similarities with the generation of online content by generative AI, particularly when protected works – or better its parts – are reproduced. Moreover, it is also likely that the CJEU decision in *SGAE*²⁹, which pertained to the transmission of TV programs to individuals in hotel rooms and determined that providing the program to individuals constitutes an act of communication to a new public, is applicable in this context as well. Given the described risks for AI developers in relying on the private copying exception, looking into the derivative works concept is advisable.

A technical challenge is the difficulty to distinguish between protected and unprotected works by automated means. Obviously, generative AI can produce output that resembles unprotected works³⁰ which does not infringe on any IP right. This needs to be emphasized, considering the uncertainties regarding the proliferation of data property concepts.³¹ However, these concerns with "regulation of AI via the allocation of property rights"³² is real as far as many data sets contain copyright protected works or protected databases. Again, Recital 9 of the CDSMD explicitly allows TDM on unprotected works. However, when it comes to identifying these unprotected works, data collectors still encounter the unsurmountable challenge posed by the low threshold of originality required for a work to be protected. Originality is now accepted for portrait photography³³, eleven consecutive words³⁴, and non-original databases. Countries like Germany and Austria additionally provide related rights for photography disregarding the requirement of originality altogether.³⁵ The issue becomes particularly

²⁵ CJEU Case C-435/12 *ACI Adam BV and Others v Stichting de ThuisKopie, Stichting Onderhandeligen ThuisKopie vergoeding* [2014] ECLI:EU:C:2014:254.

²⁶ CJEU Case C-160/15 *GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker* [2016] ECLI:EU:C:2016:644 [55].

²⁷ Compare the complaint *Getty Images (US), Inc. v. Stability AI, Inc.*, 1:23-cv-00135, p. 2.

²⁸ CJEU Case C-265/16 *VCAST Limited v RTI SpA* [2017] ECLI:EU:C:2017:913.

²⁹ CJEU Case C-306/05 *Sociedad General de Autores y Editores de España (SGAE) v Rafael Hoteles SA* [2006] ECLI:EU:C:2006:764.

³⁰ The field of unprotected works includes works and databases which are in the public domain due to various reasons, such as the absence of the necessary characteristics for legal protection, the expiration of copyright or database rights, or the deliberate release as open source. Cf. Andres Guadamuz, *Copyright infringement in artificial intelligence art*, TechnoLlama blog (2022), available from www.technollama.co.uk/copyright-infringement-in-artificial-intelligence-art.

³¹ For example, Thomas Margoni & Martin Kretschmer, *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology*, 687 *GRUR International*, 71, 8 (2022).

³² Thomas Margoni & Martin Kretschmer, *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology*, 688 *GRUR International*, 71, 8 (2022).

³³ In case CJEU C-145/10 *Eva-Maria Painer v Standard VerlagsGmbH, Axel Springer AG, Süddeutsche Zeitung GmbH* [2011] ECLI:EU:C:2011:798 the court extended copyright protection to portrait photography whilst providing in general that a "photograph [must be] an intellectual creation of the author reflecting his personality and expressing his free and creative choices in the production of that photograph".

³⁴ CJEU Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465 [61, 64].

³⁵ § 74 Austrian Copyright Act, § 72 German Copyright Act. As a sidenote, a purely digitally produced images are currently not protectable as photographs in Germany following the decision 2 U 12/16 *Kart* by the Kammergericht Berlin.

problematic when the opt-out is not feasible for the rightsholder, such as in cases where the protected work is already part of a dataset that is beyond the control of the rightsholder. Given the low threshold of originality combined with the probably impossible challenge to distinguish between protected and non-protected works by automated technical means, relying on the lawful derivative concept again seems to be an interesting option.

B. Harmonizing EU Law and National Concepts

International copyright law provides that the copyright holder holds all rights regarding adaptations and consequently their communication or making available to the public without providing any express exceptions (cf. Art. 2(3), 12 Berne Convention). In parallel, EU copyright law does neither provide a coherent horizontal definition of derivative works nor provide any exceptions.³⁶ Art. 5(b) Database Directive³⁷ and Art. 4(b) Computer Programs Directive³⁸ merely mention adaptations as restricted acts. However, it is largely agreed upon that EU law regulates derivative works indirectly via the scope of the reproduction right.³⁹ However, given the absence of a consistent EU framework for derivative works, it is imperative to consider national concepts. *Senftleben* suggests that common requirements for derivative works can be derived from fair use concepts in national laws. They often rely on “sufficient distance from protected parts of the original work” and “an assessment of the degree of originality of the prior work on the one hand, and the degree of originality of the derivative work on the other”.⁴⁰ Consequently, we will have to ask case by case (1) whether there is sufficient distance to the original and (2) if the derivative work even represents an original creation. In the following, I will juxtapose these two criteria with the requirements established by the CJEU to define a working concept for derivatives and also provide a personal view on how the assessment could be updated to better fit generative AI.

C. Distance to Protected Works

First, we will address the distance requirement. Given that the central issue with generative AI is the reproduction of parts of original works, we should initially reflect on this aspect. In the view of the CJEU, the reproduction of even a small part – such as eleven words from an article⁴¹ – may constitute an infringing reproduction. Regarding a rather small part of a musical recording, the court held in *Pelham* that the “modified form [must be] unrecognizable to the ear”.⁴² As a result, the court not only holds a low threshold for originality but also maintains a rigorous stance on protecting parts based.

It is common that generative AI models reproduce protected parts in a manner recognizable to the

³⁶ It is a question of an ongoing debate, cf. Eleonora Rosati, The right of adaptation has not been generally harmonised at the EU level: true or false? The IPKat, 1.5.2024, available from <https://ipkitten.blogspot.com/2014/05/the-right-of-adaptation-has-not-been.html>.

³⁷ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

³⁸ Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs.

³⁹ Martin Senftleben, Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU, *Pelham*, 751 *International Review of Intellectual Property and Competition Law* 51 (2020), 751.

⁴⁰ *Ibid.*, 753.

⁴¹ CJEU Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465 [61, 64].

⁴² CJEU Case C-476/17 *Pelham GmbH and Others v Ralf Hütter and Florian Schneider-Esleben* [2019] ECLI:EU:C:2019:624 [39].

average consumer. To lower the risk of such infringement, developers should therefore implement training methodology that refrains from generating output with substantial resemblances to the training data or deletes such output before publication.⁴³ It should be noted, that the same challenge arises in regard of personal data reproduced in the output.⁴⁴ Developing a common technical standard would at least reduce the risk of providers of being held liable for intentional infringement.⁴⁵ However, the strict recognizability requirement can also be criticized. It certainly does not appropriately take creators freedom of expression or any *de minimis* logic into account.⁴⁶ To mitigate its strictness, Senftleben proposes to first take the average consumer's sensibility as the reference point for recognizability. In the context of "sound mosaics," he then emphasizes that a high degree of compositional complexity can result in individual components being unrecognizable in a wider sense.⁴⁷ Applying his mosaic-approach to generative AI output, it could be argued that output remains lawful as long as the original parts largely blend into the broader composition – which is typical for output based on various data sources.

A different question is whether the appropriation of genre or personal style can result in an infringement. In general, copyright protection is only triggered by specifically expressed features of a subject, like distinctive characters in a narrative or specific sentences, melodies, or images. On the contrary, some argue that the style of an entire genre cannot be protected due to the emphasis on the abstract idea under the idea-expression dichotomy.⁴⁸ The closest the CJEU came to ruling on the issue was in its *Cofemel* decision. The court argued that a "specific, aesthetically significant visual effect" cannot be protected due to a lack of "sufficient precision and objectivity".⁴⁹ However, it is conceivable that some elements of a style could still hold enough detail and thus resemble a protectable expression. However, granting individuals IP rights of a style or genre including particular elements would certainly jeopardize the ability of others to work in the same artistic field.⁵⁰ Moreover, it is probably often the case that particular stylistic elements were originally introduced into a (sub-)culture without the author claiming any copyright protection, thus granting it to the public domain. In summary, the protection of general styles does not align with the modern approach of balancing the interests of various stakeholders in an open and creative marketplace. However, the situation is different when we speak about the personal style of an artist or – less personal – a production company. In the pursuit of coherence, it is advisable to again apply the concept of derivative works.⁵¹ The test could assess whether a derivative work maintains a sufficient degree of distinction from the

⁴³ See, for example, the approach taken by Nikhil Vyase & Sham Kakade & Boaz Barak, Provable Copyright Protection for Generative Models, 21.2.2023, archiv:2302.10870v1; see also Andres Guadamuz, Copyright infringement in artificial intelligence art, TechnoLlama blog (2022), available from www.technollama.co.uk/copyright-infringement-in-artificial-intelligence-art.

⁴⁴ For more technical details of such processes see Antonio A. Ginart, Making AI Forget You: Data Deletion in Machine Learning, arXiv:1907.05012 (2019).

⁴⁵ For a more comprehensive understanding of the scope of damages awarded in cases of intent compared to negligence see Recs. 17 and 26 of the Enforcement Directive 2004/48/EC, see also Art. 45 TRIPS Agreement.

⁴⁶ Martin Senftleben, Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU, Pelham, 759 International Review of Intellectual Property and Competition Law 51 (2020).

⁴⁷ Compare *ibid*, 757 et seq.

⁴⁸ Andres Guadamuz, Copyright infringement in artificial intelligence art, TechnoLlama blog (2022), available from www.technollama.co.uk/copyright-infringement-in-artificial-intelligence-art.

⁴⁹ CJEU Case C-683/17 *Cofemel* – Sociedade de Vestuário SA v G-Star Raw CV [2019] ECLI:EU:C:2019:721 [53].

⁵⁰ Compare for US law Ryan Abbott, Research Handbook on Intellectual Property and Artificial Intelligence (2022) 68; Stephen Wolfson, The Complex World of Style, Copyright, and Generative AI, Creative Commons Blog, 23.3.2023, available from <https://creativecommons.org/2023/03/23/the-complex-world-of-style-copyright-and-generative-ai/>.

⁵¹ William F. Patry, Patry on Copyright (2021) § 4:14 remains ambivalent on the issue in the US law context.

aggregated original works that collectively contribute to the overall personal style.⁵² If the derivative work replicates a substantial level of detail, it may be deemed as infringing. The use of specific AI training methodologies could again mitigate such situations. Nevertheless, beyond employing the derivative test, personal styles should not be protected by copyright. We should follow the precision and objectivity requirements of the CJEU to avoid impeding creative innovation and education.

Generative AI is certainly not the first technology that enables the reproduction of a specific style of an artist. The widespread adoption of the internet and even preceding modern technologies already enabled reproduction en masse. However, AI models do not necessarily increase mass reproduction but enable the appropriation of the style itself. Generative AI thereby touches upon a central commercial interest of creatives – the ability to produce custom works specific to a certain context (commissions) which has for centuries been a central source of income for creatives. Impeding on this established interest, undertakings will probably increasingly appropriate individual styles and proliferate generated content. US artists' unions are currently engaged in an enduring strike, arguing the threat posed to their economic well-being by advancements in AI.⁵³ Therefore, whilst fully protecting personal style by copyright could lead to stifling overprotection, establishing a fair remuneration system could strike the balance. Some form of remuneration is certainly in accordance with copyright principals. Commercial undertakings offering generative AI services should be required to provide fair remuneration for the use of distinct personal style. Since prompting is commonly done in human language, it is also easy to prove that a personal style has been appropriated.

D. Originality of the Derivative Work

Beyond the distance criterium, national concepts often additionally require originality of the derivative work. However, the low standard of originality under EU Copyright law already tilts the weight towards the element of distance compared to the one of originality. In particular, regarding works protected by neighboring rights, such as non-original photography or phonogram recordings, the notion that a derivative work needs to be original in itself seems unsound.⁵⁴ Moreover, if we think about originality and its etymological origin, it is not completely evident that it necessitates a human author. Originality primarily pertains to a work being a novel expression.⁵⁵ The CJEU's generally holds that originality is intrinsically connected to an author's own intellectual creation.⁵⁶ From this we could draw the following conclusion: Since originality is a requirement for derivative works on the national and the European level, every derivative work would necessarily need to be created by a human being and be alleageable for copyright protection by itself.

However, given the already stressed low threshold for copyright protection and the consequences of the CJEU's *Brompton* ruling, the result of the analysis could shift. There is some chance that low-level authorial and even non-authorial AI works could count as derivative works (and possibly even stand-

⁵² Cf. Ryan Abbott, Research Handbook on Intellectual Property and Artificial Intelligence (2022) 69 for US law.

⁵³ Compare The Guardian, Talks between striking Hollywood actors and studios break down, 12.10.2023, available from <https://www.theguardian.com/culture/2023/oct/12/actors-union-talks-suspended-sag-aftra-hollywood-strike>.

⁵⁴ Martin Senftleben, Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU, Pelham, 758 International Review of Intellectual Property and Competition Law 51 (2020).

⁵⁵ From greek *origō* = act, event or process of coming into existence: beginning, origination, see <https://en.wiktionary.org/wiki/origo#Latin>.

⁵⁶ CJEU Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECLI:EU:C:2009:465 [35]; confirmed by CJEU Case C-145/10 Eva-Maria Painer v Standard VerlagsGmbH, et alia [2011] ECLI:EU:C:2011:798 [87].

alone original works). In the context of the design of a folding bicycle, the CJEU's *Brompton* decision effectively lowered the threshold for originality to an individual expression integrated within predefined technical constraints. Protection may be granted when “a product whose shape is, at least in part, necessary to obtain a technical result, where that product is an original work resulting from intellectual creation, in that, through that shape, its author expresses his creative ability in an original manner by making free and creative choices in such a way that that shape reflects his personality”.⁵⁷ Given that AI output is generated based on prompts individually chosen by a human, we could therefore either argue (1) that the prompt is the author’s own intellectual creation and the output is the protected derivative⁵⁸ or (2) – a approach which avoids the fragmentation of the creative process –⁵⁹ that they form an entirety protected as a whole. This perspective would also align with the *Brompton* concept that creativity can thrive even in the presence of technical limitations. Comparable reservations once applied to photographic images. The generative AI model could then be understood as an artistic devise played by prompting and the process being eligible for copyright (or *de lege ferenda* at least deserving of a neighboring right). Since it is not mentioned once, we could also entertain that in *Pelham* the CJEU dropped the – already weak – originality criterium altogether in the context of derivatives.⁶⁰ In summary, we could either argue for the protection of generative AI output based on creative prompting or drop the originality requirement for the confined context of derivatives. Derivative works would then mostly be tested for recognizability (distance) as the main element of the test – an element that is somewhat manageable by means of automated technology. In alignment with the EU derivative concept, recognizability would then delineate the boundaries of the reproduction right rather than legitimizing exceptional infringement.

Whilst I cannot go into details with other potentially useful exceptions⁶¹ allow me to present my wider perspective of the issue at hand. At first glance, it seems that we are faced with a choice: either protect originality in the face of automation or give it up for a somewhat flat mechanism comparing works with each other. Given the current trajectory towards a creative world heavily reliant on automated assistance, however, it might be practical to consider embracing the latter option. However, we could also explore a new dual-track system that divides the derivative test into two distinct pathways. On the first track, the recognizability criterium would mainly apply to derivative works including automated output. On the second pathway, genuine human artistic originality could legitimize more severe infringement. This approach would allow us to make the necessary differentiation between automated commercial reproduction and authentic human artistic expression.

⁵⁷ CJEU Case C-833/18 SI, *Brompton Bicycle Ltd v Chedech/Get2Get* [2020] ECLI:EU:C:2020:461 [38].

⁵⁸ This argument was entertained by Thomas Margoni in a Webinar in 2023.

⁵⁹ We would probably not consider the brushstroke and the painting as separate acts for the purpose of copyright, at least, as long as there is only one author.

⁶⁰ Martin Senftleben, *Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU Pelham*, *International Review of Intellectual Property and Competition Law* 51 (2020), 759 et seq.

⁶¹ The traditional exceptions for quotation, criticism, review, caricature, parody and pastiche (cf. Art. 17(7) CDSMD) represent a whole different universe of lawful derivative works. Since I cannot go into the details here, I will only provide one remark for further consideration. In the course of testing whether the exceptions applies, human prompts allow us to determine the intention of the user. Combined with the degree to which the desired theme is expressed in the final output, strong arguments either for or against the fulfillment of the exception can be brought forward. Lastly, the pastiche exception in the CDSMD could hold much potential for generative AI. One might speculate, that it could override the recognizability criterium altogether.

VI. CONCLUSION

Since generative AI represents a data analysis process that results in the generation of information, it fits under the flexible concept of Arts. 3 and 4 CDSMD. We also find anchor points in the Directive that enable the development of AI models which generate marketing or entertainment content. Statistical residue data within the model and generated output also represent typical TDM (sub-)processes and should therefore fall under the exception. However, these acts are only permitted for internal use by the developer (R&D). Moreover, we have seen that the exceptions in Art. 5 (1) Information Society Directive and Arts. 6 and 9 Database Directive do not support the generative AI development process.

Regarding the provision of generated output to third parties (commercial deployment), the TDM exception does instead not apply. Whether generated output can alternatively count as lawful derivative work will depend on further clarification of the concept by the EU legislator or the CJEU. By focusing on the low threshold for originality and the recognizability criterium, the CJEU should allow non-authorial output to fall under the derivative works concept. To be lawful, a derivative AI work that incorporates parts of original works would then require that these parts fall into the background of the generated work. In this context, personal style should not be protected by copyright but rather be considered in an adapted derivatives test. Given the large economic impact on creatives, however, generative AI services should be required to provide fair remuneration when they use distinct personal style.