Developing Artificial Intelligence-Based Content Creation: Are EU Copyright and Antitrust Law Fit for Purpose?

Juha Vesala [ORCID]

Abstract Artificial intelligence offers promising applications for content production. However, their development faces significant copyright issues because it involves reproduction of protected subject matter and requires datasets so large that obtaining licences from all rightholders is unfeasible. These issues potentially hinder technological development and content production. On the other hand, some AI applications can threaten the interests and incentives of those who create works and subject matter that are protected by related rights. This article examines whether EU copyright and antitrust law are capable of addressing these challenges. It identifies possibilities and obstacles in applying exceptions for text and data mining (TDM) and temporary copying to the development of artificial creativity (AC) applications. The article also examines mechanisms by which EU antitrust law facilitates access to copyright-protected training materials and licences – an important complement to the copyright exceptions. While copyright and antitrust law enable the development of AC in certain situations, their tools are limited to particular types of AI applications, certain categories of subject matter and specific market conditions, and are subject to requirements concerning the development process as well as considerable legal uncertainty. Copyright and antitrust law also remain largely toothless against contractual and technological restraints, while recent EU initiatives dealing with data access also provide little relief in this regard.

Keywords Artificial intelligence · Copyright · Competition law · Exceptions · Data

J. Vesala (✉)
University Lecturer, University of Lapland, Rovaniemi, Finland
e-mail: juha.vesala@ulapland.fi

1 Introduction

Artificial intelligence (AI) offers promising applications for producing content such as music, written text, images, and videos. While applications are still mostly experimental and under development, some are already available commercially. For instance, services that generate images are already offered and various AI tools that assist in content production are available.[1]

Automated production of content with AI (known as "artificial creativity" or AC) is often based on neural networks – a machine-learning technology. Neural networks can be trained with pre-existing content to model features of the training data and to produce outputs with similar features.[2] While technically attractive in many ways and more capable than conventional computational techniques, training neural networks for the purpose of creating content is likely to require reproduction of copyright-protected training materials, the licensing of which is often unfeasible due to the large number of works needed.[3] As this results in major copyright infringement risks, the ability to develop AC applications hinges on whether works and other subject matter can be used without rightholder consent under a copyright exception or on another legal basis permitting their use.[4] As using works in AC development without consent can significantly affect incentives to create content – since their outputs can infringe or otherwise freeride on the training works – copyright and other policies need to balance the benefits that AC brings to and the risks that AC poses to content production.

While copyright exceptions for text and data mining (TDM) and temporary copying have been examined in legal scholarship, their application to AC development raises distinct policy issues, as noted above, and involves novel legal issues on the application of the exceptions to AC development (e.g. what constitutes lawful access or an effective reservation), which, as a practice and owing to its context and aims, differs from TDM and the development of other AI applications. Notably, the exceptions were not enacted with AC or even AI development in mind, thus leaving various crucial questions wanting for manageable solutions. This article contributes to the existing scholarship by (1) analysing these copyright challenges in the specific context of AC development, (2) presenting solutions compatible with existing case-law that seek to strike an appropriate balance between the rights and interests affected (e.g. governance of reservations), and (3) analysing copyright exceptions and antitrust doctrines, which complement and interact with each other, in order to identify the gaps that remain in the overall legal framework that currently

---

[1] For examples *see* Sobel (2017); Sartor et al. (2018); Lemley and Casey (2021); Iglesias et al. (2021); Franceschelli and Musolesi (2021); Chiou (2020); Sartor et al. (2018).

[2] *See* Drexl et al. (2021); Chiou (2020); Sobel (2017); Sartor et al. (2018); Franceschelli and Musolesi (2021).

[3] For example, a dataset used in training certain AC applications contains over 5 billion images. *See* https://laion.ai/blog/laion-5b/.

[4] *See e.g.* Chiou (2020); Sobel (2017); Margoni (2018); Vesala and Ballardini (2019). The risks are not just theoretical: recently a class-action suit was filed claiming billions of dollars of damages for the infringing use of source code in AI development (*DOE 1 et al. v. GitHub, Inc. et al.*, 3:22-cv-06823, N.D. Cal, 13 March 2022).

determines the effective ability to access and use copyright-protected training data (e.g. technical, contractual and practical obstacles beyond copyright infringement risks).

To these ends, this article examines how and to what extent EU copyright and antitrust law currently enable the use of works in AC development and how the various rights and interests involved can be accommodated in their application. First, the article sets out the numerous ways in which training and use of neural networks can involve the reproduction of works (Section 2). Second, the article examines the conditions, scope and uncertainties regarding the applicability of key exceptions in EU copyright law to allow AC development, in particular the applicability of the exception for TDM (Art. 4 DSM Directive) and the exemption for temporary reproduction (Art. 5(1) InfoSoc Directive) (Section 3). Third, the article analyses the approaches by which EU antitrust law facilitates access to and use of training data by banning refusals to license, and abusive and restrictive licensing practices (Section 4). Fourth, boundaries and gaps in the copyright and antitrust framework with regard to accessing and using copyright-protected subject matter in AC development are identified, followed by discussion of the implications of EU legislative initiatives dealing with data access and the need for further policy intervention (Section 5). Finally, the article concludes (Section 6) by highlighting the possibilities and shortcomings of the current legal framework as regards AC development.

2 Copyright Infringement in Development of Artificial Creativity Applications

Neural networks are an essential technology underlying modern AI applications.[5] The training of neural networks involves feeding data into the input nodes of the networks, from which the data is fed into nodes in subsequent layers and ultimately into output nodes. By observing outputs generated by a neural network and adjusting the parameters of the network, a neural network can – with numerous rounds of training – learn to produce increasingly desirable results.[6] For instance, if the objective is to generate the image of a fox, the parameters of the network can be modified so as to gradually result in outputs that look more fox-like.

The use of neural networks to develop AC applications requires training materials that possess the characteristics of the content to be produced. Since materials consisting of creative content are by their nature often protected by copyright or related rights, those rights may be infringed at various stages of the training. First, works may need to reproduced when gathering and preparing data for

---

[5] Neural networks can consist of different numbers of nodes and layers. The nodes can carry out different functions and be connected in different ways. Their designs and sizes vary depending on the nature and complexity of the intended task.

[6] For technical details, *see e.g.* Drexl et al. (2019); Margoni (2018); Hartmann et al. (2021), Kop (2020); Gillotte (2020); Franceschelli and Musolesi (2021).

neural network training.[7] For example, acquiring images of foxes and converting the images to the required format is likely to entail reproduction of entire images.

Second, when copyrighted material is used to train neural networks, the data can be processed in ways that constitute reproduction.[8] When data is fed into a neural network, literal copying of training data may occur. For example, all the pixels of the image of a fox might be reproduced in the inputs at that point in time. In some cases, though, infringing reproduction at this stage could be avoided by using excerpts from works (e.g. very short extracts of text) if the reproduction of copyright-protected elements at any single point in time is avoided.[9]

Third, once data from input nodes enter the subsequent layers of a neural network, the data inputted may be processed in such complex ways that literal infringement is avoided. For instance, the individual pixels of the image of a fox may be processed with other pixels so that no protected aspects of the image can be observed within the neural network. However, once a neural network learns to reproduce features of the training data, protected aspects or elements of works used in training a neural network could also be reproduced within the neural network.[10] For instance, excerpts or protected features of training materials could be reproduced (e.g. parts or aspects of an image that enjoys copyright protection).

Finally, the outputs of neural networks when training or using the network may also constitute reproduction that infringes protected elements of the works used in training the application.[11] Whether the outputs produced infringe copyright in the training works depends on the design of the neural network and the application it is used in.[12]

3 Applicability of Copyright Exceptions to Use of Works in AC Development

As we have seen, the reproduction rights of copyright holders may be implicated at various stages of the AC development process. Importantly, it is difficult to avoid copyright-relevant reproduction entirely because – at least in the initial phases – entire works are likely to be copied, while in subsequent phases protected aspects can also be reproduced at a single point in time in such a way that the materials used

[7] For infringement at this stage, see e.g. Chiou (2020), paras. 8–9; Sobel (2017). In this respect, artificial creativity applications are akin to TDM and other machine-learning pursuits. For an overview of infringement risks in TDM see e.g. Geiger et al. (2018a, b); Schönberger (2018); Jackson (2021).

[8] As regards infringement in this regard, see e.g. Chiou (2020), para. 10.

[9] C-429/08 *Football Association Premier League and Others* EU:C:2011:63, paras. 157–159. Since some types of neural network, e.g. those known as long short-term memory (LSTM), can retain information from earlier rounds of inputs, this means that reproducing protected aspects of works could nonetheless take place within the network, as discussed next, even when feeding of inputs does not.

[10] Additionally, a neural network itself can constitute reproduction of the works used in training it. See e. g. Sobel (2017), pp. 63–65; Chiou (2020), para. 11.

[11] Sobel (2017), pp. 65–66.

[12] On designs of neural networks in this regard see e.g. Franceschelli and Musolesi (2021), pp. 6–7 (discussing how likely different training models are to infringe, and safeguards against the reproduction of, training data).

to train the neural networks are infringed – unless a licence, copyright exception or other legal basis authorises reproduction.

Unfortunately, hundreds of millions of items of protected subject matter may be required in order to train neutral networks. And these may have various rightholders and be subject to several forms of copyright protection at the same time (e.g. as works, by related rights or *sui generis* database rights). If no centralised or coordinated licensing mechanisms are available, obtaining authorisation to use such large bodies of subject matter is unfeasible, as even identifying rightholders, let along concluding licensing agreements with them, would be economically impossible. These prohibitively high transaction costs that prevent welfare-improving outcomes may justify the use of works without consent – under certain conditions.[13] Allowing the use of works without the rightholders' consent, even if authorisation were possible, can also be justified by the public benefits of promoting technical development and improvements in the production of the content that it facilitates.

Existing exceptions in EU copyright law – in particular the exemption on temporary copying and the exception for TDM – can already allow works to be used in AC development. However, to what extent they can do so stumbles upon questions about the applicability to AC development of certain conditions of the exceptions as well as technical and other obstacles in meeting their requirements. Below, the applicability of these exceptions to AC development is analysed, highlighting potential issues and developing approaches for resolving them under EU copyright law.

3.1 Exception for Text and Data Mining

Exceptions for TDM in the Directive on copyright in the Digital Single Market[14] (DSM Directive) are among the leading candidates for enabling AC development. Article 4 DSM Directive permits the reproduction of lawfully accessible materials for TDM purposes if a rightholder has not reserved that right. This exception is not limited to non-commercial research as is the case for the TDM exception in Art. 3 DSM Directive, but instead applies to any person or organisation and objective of TDM.[15] While the general TDM exception (Art. 4 DSM Directive) covers the use of

---

[13] Lemley and Casey (2021); Gordon (2003).

[14] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ EU L130/92 (17.5.2019) (DSM Directive).

[15] The exception for TDM of Art. 3 DSM Directive is limited to research organisations and cultural heritage institutions and solely non-commercial research purposes. Therefore, it does not apply to commercial AC development. It might be possible to commercially use the results of non-commercial research benefiting from the exception for non-commercial TDM. However, any foreseen commercial exploitation of results subsequently could result in the initial research objective being other than solely non-commercial by nature and thus preclude application of the exception to the research. The exception also only applies to research organisations and cultural heritage institutions and TDM carried out for research purposes, but not to any subsequent reproduction by others. Thus, if copies from training works are embedded in a trained neural network or research results, their reproduction when commercially developing or using the developed neural network could infringe the works.

protected subject matter in AC development, some of its conditions impose limitations or result in other obstacles or challenges affecting AC development, as discussed below.[16]

3.1.1 Concept of TDM

The first source of uncertainty concerns the heart of the exception – the concept of TDM itself. TDM is defined in the Directive as activities aimed at generating information such as patterns, trends and correlations.[17] This very general definition – virtually anything could be characterised as having the purpose of generating information – covers AC development in the sense that AC development entails analysing the training data in order to produce information (e.g. parameters of neural networks or outputs produced with them that capture patterns and correlations in the training data).[18] Yet the definition of TDM cannot be without limits and cover all or any use of works carried out through automated processing.[19] Furthermore, TDM is not a synonym for AI, though the concepts both overlap and diverge in terms of their objectives and techniques. For instance, the Commission proposal for the DSM Directive reflects an understanding of TDM as a tool for analysing data (e.g. scholarly literature) to glean information therefrom, but does not refer to the use of works in developing AI.[20]

However, a narrow interpretation of TDM of the kind given above is not warranted. First, the very broad definition of TDM and the examples provided in the Directive are clearly not constrained by such an understanding of TDM. Second, the three-step test (Art. 5(3) InfoSoc Directive) requires that no harm be done to the normal exploitation and legitimate interests of copyright holders.[21] This makes it unnecessary to limit the concept of TDM itself, and a narrow understanding of TDM would also preclude applications that produce societal benefits while not harming rightholders at all.[22] Finally, rightholders can in any event prevent the exception

---

[16] For an extensive analysis of the TDM exceptions, *see e.g.* Ducato and Strowel (2021); Griffiths et al. (2023), pp. 25–30. Implementation of the DSM Directive is still ongoing in many Member States. Most Member States, such as Estonia, Germany, Ireland, and the Netherlands implement Art. 4 DSM Directive almost verbatim. However, some implementations or proposals seek to clarify or otherwise interpret the ambiguous concepts relating to TDM in the Directive, as mentioned below where applicable.

[17] According to Art. 2(2) DSM Directive, "'text and data mining' means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations".

[18] For example, Ørstavik I (2022), p. 286, notes this concern.

[19] Ultimately, the three-step test requires that exceptions be specific. On the test, *see* Sect. 3.3 below.

[20] The proposal, COM (2016) 593, contains no mention of AI or machine-learning technologies.

[21] Article 7(2) DSM Directive.

[22] On the three-step test see Sect. 3.3 below.

from applying by making a reservation, thus protecting their interests and incentives, as discussed next.

3.1.2 Rightholder Reservations Preventing Application of the Exception

The TDM exception in Art. 4 DSM Directive does not apply when a rightholder expressly reserves use of its works or other subject matter for TDM by technical or otherwise appropriate means.[23] If such a reservation has been validly made, then use of those materials is not permitted under the exception for AC development.[24] If such reservations are made by a significant number of rightholders, the possibility of using works for AC development becomes correspondingly limited.[25]

Regardless of whether or not such a reservation has been made, it can be difficult to ensure that no *legally effective* reservation has been made when large amounts of training materials are used.[26] For example, materials (e.g. pre-prepared datasets) are not necessarily accompanied by such a reservation but reservations may still be made in respect of material that is lawfully accessible elsewhere (e.g. at the source where the subject matter was obtained for the dataset). Article 4 DSM does not specify where such a reservation needs to be made in order to be effective – only that it be express and made in an appropriate manner. Even if reservations are made in standardised, machine-readable form, this issue still arises because a valid reservation may have been made somewhere other than where the mined materials were obtained. Doubts over the existence of valid reservations also arise because online services use language in their terms of use that may or may not constitute an effective reservation (e.g. banning reverse engineering or similar methods, or the storing of available content) and it can also be unclear whether reservations have been made by rightholders themselves or at their behest, or only by a service provider (in which case they would not prevent mining).

These legal uncertainties, which make it difficult to automatically or even manually ensure that effective reservations have not been made, give rise to the risk that the TDM exception may not cover the use of some subject matter for the

---

[23] Article 4(3) DSM Directive.

[24] Arguably, reservations become effective from the moment they are made so as to prevent further mining under the exception, not retrospectively affecting materials already mined or the retention thereof. See e.g. German government bill to adapt copyright law to the requirements of the digital single market (19/27426, 9 March 2021), 89 (ex nunc effect). Moreover, according to Art. 4(2) DSM Directive, copies made may be retained for as long as necessary for the purposes of mining. Because it may be impossible to untrain neural networks with respect to certain training data, it should remain an option to retain copies that the neural network itself potentially constitutes until the network can be retrained with new materials from which the reserved materials have been removed.

[25] On the obstacles, *see* Chiou (2020), paras. 25–26. It is possible that rightholders will routinely begin reserving use of their works in TDM as there is little harm in doing so. According to Ducato and Strowel (2021), many platforms are already doing so in their end user agreements.

[26] On these concerns, *see* Chiou (2020), para. 25.

development of AC applications. Interestingly, arrangements are emerging in which an AI developer or dataset provider takes note of objections by rightholders to the inclusion of their works in the datasets concerned.[27] These kinds of model can provide a workable solution to governing rightholder reservations more generally in the context of the TDM exception. Not only do they provide legal certainty to those carrying out TDM – as they receive information on works that cannot be used under the TDM exception – but such models would also allow more discrete tailoring of reservations to target a specific AI developer or dataset, rather than reservations at source or elsewhere that prevent any TDM using the materials. This, though, would require that not all reservations made by rightholders elsewhere (e.g. at other sources of works or unconnected with sources, or not in machine-readable form) be deemed appropriately made if those engaging in TDM cannot automatically or otherwise reasonably identify them. Such a model would also require those seeking to benefit from the TDM exception to divulge which materials or sources of materials they are using in order for rightholders to be able to make reservations concerning their works.[28]

3.1.3 Lawful Accessibility of Works

A related source of uncertainty is the prerequisite that the exception only applies to works and other subject matter that are lawfully accessible. The concept of *lawful accessibility* does not directly correspond to existing autonomous terms of EU law, such as lawful use, and is not defined in the DSM Directive.[29] While explicitly licensed sources (e.g. subscription databases) generally meet this requirement, there are materials whose status is unclear in this regard. A fundamental problem with the concept is that it refers to an aspect – access – that copyright addresses only to a certain extent because exclusive rights mostly apply to providing access or reproduction, not accessing works as such. To what extent circumvention of technical measures, violation of conditions of a licensing or other agreement or avoidance of geo-blocking measures taint lawful access under the DSM Directive is far from clear – in each case an aspect of the process of accessing works may be unlawful, but it is not apparent that this is intended to preclude application of the TDM exception.

Even though access to works is often lawful under copyright law, uncertainty about the substance of this requirement can significantly hamper AC development.

---

[27] For example, a major AC developer is developing an opt-out system (https://stability.ai/faq) and has taken other measures to protect authors (https://discord.com/channels/1002292111942635562/1002292112739549196/1045335271119523961). One content-sharing service also allows its users (creators of content) to opt out from use of their images in AI development and applies machine-readable tags to their website to this effect for third parties to take note of this fact. It has also modified its terms of use to require compliance with the opt-outs (https://www.deviantart.com/team/journal/UPDATE-All-Deviations-Are-Opted-Out-of-AI-Datasets-934500371).

[28] One initiative (https://haveibeentrained.com/) allows people to determine whether a particular image is included in a popular dataset used in AC development.

[29] Also, the examples mentioned in Recitals 10 and 14 of the Directive are diverse: they include licensed sources, but also mention content freely available online. For discussion of this and related concepts in EU copyright, *see* Synodinou (2019).

For example, some EU Member States have considered that lawful access is precluded if the copy to be used for further reproduction to carry out TDM has been reproduced or made available to the public in an infringing manner, or technical measures have been circumvented.[30] This understanding poses challenges to AI developers because verifying the origins of source materials is impossible where large datasets are concerned. Such a requirement would effectively limit mining to materials available from licensed sources, since information on the copyright status of source works tends not to be publicly available. The approach is also problematic conceptually as it presupposes that the source copy to be used for TDM has a determinable copyright status (infringing in its origin or resulting from unlawful circumvention), which is often not true.[31] A more workable and conceptually justifiable approach is to require a sufficient duty of care from those engaging in TDM – taking into account the nature of sources for the data used (e.g. reputable websites, content-sharing platforms, pirate websites) and other factors when assessing whether or not the mined subject matter could be deemed to be lawfully accessible.

3.1.4 Applicability of the TDM Exception to AC Development

The TDM exception (Art. 4 DSM Directive) allows AC development in a wide variety of circumstances and covers a wide range of works and related rights, including press publishers' rights and sui generis database rights. However, the TDM exception does not apply if rightholders have reserved the right to mining, the absence of which is difficult to ascertain in the case of large datasets and because the Directive does not limit where or for whose benefit effective reservations can be made. These uncertainties can be alleviated – when interpreting exceptions – by recognising and supporting the organisational models that are starting to emerge for governing reservations. Moreover, as long as there is no understanding of what subject matter is "lawfully accessible", AC developers face uncertainty because some interpretations would require examining the origins of the works and subject matter to be used in AC development in terms of their status under copyright law. Instead, this condition should be understood as merely requiring that the miner could deem the content lawfully accessible given the source and other factors.

---

[30] *E.g.* the Finnish Government proposal implementing the DSM Directive (43/2022) explicitly provides for such an approach. A Swedish Ministry Memorandum (DS 2021:30), pp. 63–64, also leans towards such an approach.

[31] *E.g.* different laws may apply to the act of making available, reproducing, or circumventing technical measures and the act of TDM – different conclusions are possible depending on the perspective from which the question is examined. Similar dualities also arise when copyright limitations or rules on limited intermediary liability apply so that the source copy has been lawfully made and made available from one perspective (*e.g.* for research purposes) but the same or related acts can also be infringing from another perspective (*e.g.* for commercial purposes). An emerging issue is whether access to data mandated by other legislation confers lawful access under the TDM exception. On these access mechanisms, *see* Sect. 5.2 below.

3.2 Exemption for Temporary Reproduction

Another key provision in EU copyright law that can cover AC development is the exemption on temporary copying (Art. 5(1) InfoSoc Directive).[32] The exemption applies when reproduction is temporary and transient or incidental, occurs as part of a technical process, is carried out solely for the purpose of lawful use, and copying has no independent economic significance.[33]

The temporary copying exemption was not enacted with this type of copying in mind. It was rather intended to address temporary copying when consuming, accessing and transmitting works – not copying for the purpose of producing content.[34] However, according to the CJEU, the objective of Art. 5(1) InfoSoc Directive is to enable technological development and the use of technologies while safeguarding a fair balance between the interests of rightholders and users of works.[35] As argued below, the temporary copying exemption can be applied to AC development in line with this objective of enabling the development and use of technologies in a balanced manner and otherwise in compliance with how the CJEU has construed the conditions of the exemption.

3.2.1 Reproduction as an Integral Part of the Technological Process

For the temporary copying exemption to apply, copying must first constitute an integral part of a technological process. This entails that copying does not occur outside of the process and that copying is necessary for the correct and efficient functioning of the process.[36]

This condition does not pose an obstacle to AC development as the copying of training data that takes place when developing and using neural networks can be necessary for the correct and efficient training and use of a neural network.[37]

---

[32] 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 (InfoSoc Directive) [2001] OJ EU L167/10 (22.6.2021), Art. 5(1). On the relevant case-law, *see e.g.* C-5/08 *Infopaq International A/S v. Danske Dagblades Forening* EU:C:2009:465; C-360/13 *Public Relations Consultants Association Ltd v. Newspaper Licensing Agency Ltd and Others* EU:C:2014:1195; C-403/08 and C-429/08 *Football Association Premier League and Others* EU:C:2011:63; C-302/10 *Infopaq International v. Danske Dagblades Forening* EU:C:2012:16 (Order). The optional exceptions in the InfoSoc Directive that Member States can adopt are either limited to non-commercial activities or cover commercial development of artificial creativity applications only to a limited extent, or apply only in a specific economic domain, such as press reporting, and apply only in Member States that have adopted the exception in their national legislation. On applicability of the optional exceptions specifically to AI-based news production, *see e.g.* Vesala (2022). For applicability of the mandatory and optional exceptions to TDM, *see e.g.* Triaille et al. (2014); Geiger et al. (2018).

[33] For an overview, *see* Triaille et al. (2014), 41–45.

[34] *See e.g.* Guibault et al. (2007), p. 31; Chiou (2020), para. 14.

[35] C-403/08 and C-429/08 *Football Association Premier League and Others* EU:C:2011:631, paras. 164 and 179. Moreover, the CJEU has subsequently held that temporary copying for producing summaries of news articles could be exempted, which in turn suggests that the exemption can apply to production of content. C-5/08 *Infopaq International A/S v. Danske Dagblades Forening* EU:C:2009:465.

[36] C-302/10 *Infopaq International* EU:C:2012:16, para. 30.

[37] *See also* Chiou (2020), para. 15; Margoni (2018), p. 17.

Training and using neural networks undoubtedly constitutes a technical process, which could not be carried out at all or as effectively without repeatedly copying the training data. Moreover, copying fewer times or to a lesser extent may result in inferior performance of the neural network and a lower likelihood of successful training. AC development can thus meet this requirement if the training data is not used for other purposes and the process is designed appropriately.

3.2.2 Temporary and Transient/Incidental Nature of Copying

Reproduction as an integral part of a technological process further needs to be temporary and either transient or incidental. Copying is considered *temporary* when copies are later deleted either by a human or automatically.[38] It is *transient* if copies are not retained longer than necessary for the purposes of the technological process but are automatically deleted as soon as possible without human intervention.[39] And it is *incidental* if it has no purpose other than being a part of the technological process, that is, even if retained longer.[40]

Reproduction that occurs during training of neural networks can satisfy the criteria of being temporary as well as transient or incidental as understood in CJEU case-law. Typically, the data fed into a neural network is deleted almost immediately once it passes the network, and is not stored.[41] Copies made during training may therefore qualify as both temporary and transient.

However, other copies made during the development and use of a neural network are not necessarily automatically deleted or removed manually. For instance, when training materials are collected for the purposes of training neural networks, copies of works in the training dataset are not automatically erased and hence are not transient. Moreover, the trained neural network itself could be considered a non-transient copy of the training works if it allows the reproduction of works used in its training, similar to a compressed file.[42]

However, copies that are not transient because they are not automatically deleted can still be incidental – the alternative to being transient. This requires that the copies have not been made for purposes other than the technical process. Since datasets and neural networks may be technically necessary for the development and use of a neural network, they can be incidental as long as the copies are not used for other purposes. However, the question may arise whether incidental copies are temporary (as further required) if they are maintained without the intention of being deleted automatically or by humans. This problem could possibly be avoided by

---

[38] C-360/13 *Public Relations Consultants Association Ltd* EU:C:2014:1195, paras. 26–27. Automatic deletion is necessary to provide legal certainty to copyright holders that deletion is not up to human action. C-429/08 *Football Association Premier League and Others* EU:C:2011:631, para. 62.

[39] C-360/13 *Public Relations Consultants Association Ltd* EU:C:2014:1195, para. 40.

[40] C-360/13 *Public Relations Consultants Association Ltd* EU:C:2014:1195, para. 43.

[41] Chiou (2020), para. 10; Schönberger (2018), p. 169; Margoni (2018), p. 17.

[42] Chiou (2020), para. 16; Margoni (2018), p. 18.

regularly revising datasets so that no work is retained permanently in the dataset (to ensure that the dataset is non-temporary).[43]

3.2.3 Sole Purpose of Lawful Use

A further requirement of the exemption is that the sole purpose of temporary copying is "lawful use".[44] Lawful use is use authorised by the rightholder or not restricted by the applicable legislation (in particular, applicable EU and national copyright law).[45]

This requirement can be met where the sole objective is to develop an AC application that does not infringe copyright.[46] This would especially be the case when the outputs of the AC application do not infringe the training materials by reproducing their protected features or communicating them to the public (or if reproducing them by virtue of benefiting from another exception). In this regard, the CJEU has for instance considered that preparing summaries of news articles, non-infringing under national law, can constitute lawful use.[47] Although it may be difficult to entirely avoid some outputs of AC applications containing protected aspects from training data, it arguably suffices that the sole *purpose* of the activity is lawful use; it is not required that unintentional infringement is wholly avoided.

3.2.4 No Independent Economic Significance

The final requirement of the temporary copying exemption is that temporary copying should not have any independent economic significance. This entails that temporary copying must not produce economic value distinct from the value of lawful use of the works.[48] Profiting from providing access to temporarily copied

---

[43] And, additionally, designing neural networks so that they do not represent copies of works used in training as a result of being able to reproduce those works (in order to prevent the network from constituting a non-temporary copy). *See also* Margoni (2018); Chiou (2020), para. 16 (not incidental, as not temporary).

[44] Copying for the sole purpose of transmission is also acceptable under Art. 5(1) InfoSoc but not relevant in this context.

[45] C-527/15 *Stichting Brein* EU:C:2017:300, para. 65; C-403/08 and C-429/08 *Football Association Premier League and Others* EU:C:2011:631, para. 168; C-302/10 *Infopaq International* EU:C:2012:16, para. 42. As Margoni (2018), p. 18, notes, it is unclear whether contractual restrictions preclude lawful use.

[46] Apart from the above considerations, the CJEU has also – under the lawful use criterion – examined whether use is compatible with the three-step test, including how it affects the normal exploitation of works or legitimate interests of authors. C-527/15 *Stichting Brein* EU:C:2017:300, para. 70. This approach limits the scope of lawful use to situations where harm to rightholders is avoided. However, since Art. 5(5) InfoSoc Directive in any event requires the three-step test to be considered, as explained below, doing so separately under Art. 5(1) appears unnecessary and renders the lawful use criterion less predictable since the legal nature of the activity (infringing or not) would not be decisive, but the consequences of applying the exemption also need to be examined. Regardless, applying exceptions to AC development can be compatible with the three-step test under certain conditions, as discussed below, particularly where harm to copyright-holder incentives can be excluded.

[47] C-302/10 *Infopaq International* EU:C:2012:16, paras. 40–46.

[48] C-302/10 *Infopaq International* EU:C:2012:16, para. 50 ("However, those acts must not have independent economic significance, in that the economic advantage derived from their implementation must not be either distinct or separable from the economic advantage derived from the lawful use of the

works prevents this condition from being met, as does temporary copying that results in changes to the subject matter and thus allows the works to be exploited in a different form.[49] For example, where the consumption of works does not infringe copyright and thus constitutes lawful use, temporary copying that is limited to realising that same lawful use can satisfy the requirements of copying having no independent economic significance.

This condition has been noted as potentially posing an obstacle to TDM as independent economic value can be created as a result of the information and knowledge produced in the process.[50] In a similar vein, it could be argued that training neural networks for AI development purposes could create independent economic value owing to the valuable capabilities attained by the neural network as a result of the temporary copying of works, which would thus prevent the exemption from applying due to copying having independent economic significance.

However, CJEU case-law also allows a different conclusion, namely that the essence of the CJEU's reasoning on the requirement of copying having no independent economic significance is that temporary copying should not represent a means of profiting by providing access to temporary copies or otherwise exploiting works beyond the lawful use concerned.[51] This logic rules out temporary copying that constitutes the exploitation of works by way of providing access to copies in the same or a modified form (e.g. a different language), as this involves exploitation of works in a way not authorised by the rightholder or permitted by law (beyond the lawful use concerned).

Nothing in the case-law, though, requires that value could not be created if it does not involve exploiting copied works in the same or a modified form beyond lawful use.[52] To provide an extreme example, using the heat generated by computers that are processing works to keep a building warm may be economically valuable while still wholly separate from any value based on exploiting the works. Similarly, when trained neural networks do not embody any protected aspects of the works but only represent more abstract or otherwise non-infringing features of the works, the economic value created does not stem from exploiting the protected aspects of the works in any way. For example, if the value of trained neural networks is based on functionalities relating to English grammar, and the networks are not capable of reproducing any copyright-protected aspects of the training material, then this value

---

Footnote 48 continued

work concerned and it must not generate an additional economic advantage going beyond that derived from that use of the protected work"); C-403/08 and C-429/08 *Football Association Premier League and Others* EU:C:2011:631, paras. 174–178.

[49] C-302/10 *Infopaq International* EU:C:2012:16, paras. 52–53.

[50] *See e.g.* Geiger et al. (2018a, b), p. 11; Chiou (2020), para. 17 (questioning whether the condition can be satisfied because the collected datasets can have independent value).

[51] C-302/10 *Infopaq International* EU:C:2012:16, para. 50.

[52] When discussing transformation of subject matter, C-302/10 *Infopaq International* EU:C:2012:16, paras. 52–53, the CJEU is arguably concerned about exploitation of the subject matter in altered form beyond lawful use. For this reason, if there is no exploitation of the works (even in transformed form) or all exploitation of subject matter remains within lawful use, temporary copying of works does not amount to independent economic significance beyond the sole lawful purpose.

of the neural network created through temporary copying can be thought of to be wholly limited to the lawful use of creating such a non-infringing neural network[53] or the value to be entirely distinct from copying the works since the works are not exploited even in modified form.

Additionally, for policy reasons, preventing the creation of value not stemming from protected aspects of the works would be unwarranted. First, although it is possible that allowing profiting from copying of non-protected aspects could harm incentives to create, as will be discussed below, Art. 5(5) InfoSoc Directive in any case requires that normal exploitation of works and the legitimate interests of copyright holders not be harmed. For this reason, it is unnecessary to limit the scope of exemption already laid down in Art. 5(1) InfoSoc Directive. Second, the three-step test under Art. 5(5) InfoSoc Directive is better suited to protecting copyright-holder interests and incentives, as the test gauges the overall consequences of applying an exception, whereas the requirement of temporary copying, as it has no independent economic significance, focuses on economic benefits accruing to the person making temporary copies, not its impact on the rightholder's interest or incentives to create.[54] Finally, it would be counterproductive to preclude, under Art. 5(1) InfoSoc Directive, value creation based on copying non-protected aspects of works because that would inhibit societally valuable activities that do not harm copyright holders at all. If anything, creation of new value not derived from protected aspects of works should be treated more favourably than copying protected features of works since the former is more likely to be beneficial overall.

Accordingly, the interpretation advocated above would require that the AC application developed not produce outputs that constitute a means of exploiting works in infringing form. By contrast, if trained neural networks do produce outputs that infringe training works (e.g. automatic translations or modifications), works used in training neural networks may be exploited in a way that has independent economic value beyond the lawful use concerned, thus preventing the exemption from applying. Consequently, in addition to being intended solely to produce non-infringing outputs (under the criterion of lawful use), neural networks are to be developed to also provide such outputs with a sufficient degree of certainty in order to avoid providing an independent means of exploiting training works (to satisfy the requirement of no independent economic significance).

3.2.5 Applicability of the Temporary Copying Exemption to AC Development

The temporary copying exemption lends itself to an interpretation consistent with CJEU case-law that allows reproduction that occurs during the training of neural networks when this is carried out for the purpose of generating non-infringing outputs, provided that all copies made are automatically deleted (and datasets regularly updated if not automatically deleted) and that the outputs produced are

---

[53] Margoni (2018), p. 17 (concluding that value created can be connected to lawful use, not to temporary copies).

[54] Value to persons carrying out temporary reproduction is an inappropriate proxy for negative effects on incentives to create since copying can be entirely harmless to incentives to create and other copyright-protected interests, as such value can stem from aspects other than the qualities of the copied work.

non-infringing. However, the temporary copying exemption only applies to certain copyright-protected works (works other than computer programs and databases), subject matter protected by the related rights of performers and phonogram producers, and fixations of films and broadcasts protected by related rights.[55] This means that, for instance, content contained in copyright or sui generis protected databases cannot be used for AC development under this exemption.[56] This can be an issue in AC development when training materials have been labelled (e.g. subject or other attributes of the works) or collections of subject matter have otherwise been processed, as such efforts may give rise to database protection not covered by the temporary copying exemption. However, if the TDM exception applies to AC development, the limitation in the scope of the temporary copying exemption in this regard is not decisive, as it permits use of the databases in any event.

3.3 Implications of the Three-Step Test on the Application of the TDM and Temporary Copying Exceptions

Individually and in conjunction with each other, the TDM and temporary copying exceptions cover the use of protected subject matter for AC development under the circumstances noted above. However, application of the exceptions by a national court further requires that the three-step test be satisfied.[57] The test requires that exceptions "shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject matter and do not unreasonably prejudice the legitimate interests of the rightholder."[58] The test is relevant in the context of AC development because applying exceptions to the development of AC

---

[55] Articles 1(2), 2 and 5(1) InfoSoc Directive. However, the exemption also applies to press publications under Art. 15(3) DSM Directive.

[56] *E.g.* existing datasets or other collections of works could constitute protected databases. *See e.g.* Drexl et al. 2021, 6. Using their content may infringe the copyright or sui generis right to databases. *See e.g.* C-762/19 SIA '*CV-Online Latvia' SIA v, 'Melons' SIA* EU:C:2021:434 (obtaining and presenting data from a freely accessible database may infringe the sui generis right where it threatens investment in the database); C-202/12 *Innoweb* EU:C:2013:850 (highlighting the significance of impact on investment). However, to what extent rights (Arts. 5 and 7 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive) [1996] OJ EU L77/20 (27.3.1996)) or exceptions or users' rights (Arts. 6, 8 and 9 Database Directive) pertaining to databases restrict or allow the use of databases in AC development is an issue that goes beyond the scope of this article and requires further analysis. For analysis of AI development issues related to database protection, *see* Ørstavik I (2022) (discussing different uses of databases in AI development through this lens); Meys (2021).

[57] The exceptions in EU law are to be strictly interpreted in the light of the three-step test. C-527/15 *Stichting Brein* EU:C:2017:300, paras. 62 and 63; C-5/08 *Infopaq International* EU:C:2009:465 para. 59. In earlier decisions the CJEU concluded that meeting the requirements of the exemption of Art. 5(1) InfoSoc Directive would lead to compliance with the three-step test. C-302/10 *Infopaq International* EU:C:2012:16, paras. 56–57. However, subsequently, the CJEU has considered the three-step test separately from assessment under Art. 5(1) InfoSoc Directive without making such an assumption. C-527/15 *Stichting Brein* EU:C:2017:300, paras. 70–72; C-360/13 *Public Relations Consultants Association Ltd* EU:C:2014:1195, paras. 54–63.

[58] Dir 2001/29/EC, Art. 5(5). On the three-step test in the context of TDM, *see* Gervais (2019), paras. 50–61, Senftleben (2022) and more generally in international intellectual property law Geiger et al. (2014), Senftleben (2004).

applications may in particular contravene the requirements that the normal exploitation of works not be hindered and the legitimate interests of rightholders not be unreasonably prejudiced. This is because the outputs of AC applications that infringe, imitate or otherwise freeride on works used to develop applications can affect both the exploitation of works and rightholder interests.[59]

Whether harm inflicted by applying an exception to AC development impinges on the normal exploitation of works or legitimate interests requires case-by-case assessment of its expected consequences because the functionalities, performance (e.g. quality of outputs) and market context of AC applications vary significantly.[60] A key factor in assessing harm to rightholder interests is the nature of the outputs that would be produced with the AC application concerned. First, where *the outputs of AC applications infringe* the materials used in training neural networks, normal exploitation of works or the legitimate interests of copyright holders are harmed in a way that can fail the three-step test and thus prevent the exception from applying.[61] If applying the exception to an AC application developed would in this manner ultimately undermine sales of the works in the content product market where they are offered, normal exploitation can be hampered and the legitimate interests of authors jeopardised.

Second, *non-infringing outputs* of AC applications could also flout the three-step test. This is a relevant issue since AC applications could be used to create non-infringing content that competes with the works used in the training thereof.[62] For example, AI could be used to mass-produce summaries of text and imitations of works that, despite not necessarily infringing copyright, harm the exploitation of works by copyright holders. As a consequence, the question whether development of these kinds of AC applications is allowed becomes pressing if effective action against such mass-produced outputs cannot be taken directly.

---

[59] The normal exploitation of works and legitimate rightholder interests could also be unjustifiably harmed when application of an exception harms the licensing of works for AC development purposes. For now, there appears not to be an empirical expectation or normative basis to be able to license works for AC development purposes when the works have not been created specifically for AC development but are intended for normal consumption by humans (*e.g.* music). However, there may be a market for licensing specifically produced training data (*e.g.* synthesised data) or prepared datasets (*e.g.* collection of labelled pieces of content) that could be undermined by application of an exception that undoes the need to obtain a licence. According to Hilty and Richter (2017), p. 7, there is a market for datasets. The three-step test can pose an obstacle to applying exceptions where the subject matter concerned consists of materials produced specifically for AI development. Moreover, as noted by Sobel (2017), pp. 76–77, gaining access to user data and content may already explain why companies offer services to consumers, forming another kind of existing licensing market.

[60] The contents of the three-step test in these respects have been discussed in the World Trade Organization's Panel Report on the United States – Sec. 110(5) US Copyright Act, WT/DS160/R (15 June 2000), which in particular gives weight both to empirical and normative aspects of normal exploitation of works, paras. 6.178–6.189, and notes that additionally the concept of legitimate interests has a normative element linked to the objective of exclusive rights, para. 6.224. For CJEU case-law addressing the contents of the requirements of normal exploitation and legitimate interests, *see* footnote 64 below.

[61] Of course, this depends on how the developed AC application is designed and used – for instance, if the outputs are not offered to the public, they are unlikely to affect the exploitation of works or interests of authors.

[62] *See also* Sobel 2021 (discussing market-encroaching uses of AI); Sartor et al. (2018), pp. 11–12 (referring to similar concerns).

Since the three-step test examines the effects of an exception – and not just the effects of the infringing activities for which the exception is sought – the test can be failed if applying an exception enables non-infringing activities subsequently.[63] To this effect, the CJEU has objected to temporary copying that enabled an arguably non-infringing subsequent activity (private viewing of content), since allowing the temporary copying that makes that possible would have harmed the normal exploitation of works and legitimate interests of copyright holders.[64]

In any event, issues with the three-step test can be avoided and the TDM and temporary copying exceptions can be applied when the outputs of an AC application being developed would not infringe the works used or otherwise freeride on copyright-holder efforts, as discussed above.[65] However, where that is not the case but an AC application would, for instance, output infringing content, national courts need to examine the likely impact of the specific AC application concerned on the normal exploitation and legitimate interests of authors.

4 Antitrust Mechanisms for Facilitating Access to Copyright-Protected Training Data

While copyright exceptions permit the use of works without copyright-holder authorisation for developing AC applications in various circumstances, as observed above, this does not cover all objectives of AC development or categories of protected subject matter, and is subject to further conditions that explicitly or otherwise limit the applicability of the exceptions. Additionally, EU copyright law does not generally prevent copyright holders from limiting access to or use of works in AC development, whether contractually, technically or otherwise. For example, even if the temporary copying exemption or the TDM exception applies, EU copyright law does not preclude technical, contractual or other restraints that prevent the exempted activities.[66]

---

[63] For comparison, in some situations EU copyright legislation sets additional requirements that the beneficiary of an exception must meet that appear to address non-infringing imitation. For example, while the decompilation of computer programs is permitted pursuant to an exception in order to achieve interoperability, the information obtained may not be used for other purposes or to create a substantially similar computer program. Nor may the exception be otherwise used in ways that conflict with the three-step test. Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, OJ EU L111/16 (5.5.2009), Art. 6(2) and Recital 15.

[64] C-527/15 *Stichting Brein* EU:C:2017:300, para. 71. Moreover, A-G Trstenjak has suggested that even non-infringing summaries of news articles could be objectionable on these grounds. C-5/08 *Infopaq*, Opinion of A-G Trstenjak, EU:C:2009:89, paras. 138 and 140.

[65] However, as noted above, the impact of applying an exception on licensing markets also needs to be considered, as the use of works specifically created for AC development purposes in particular can contravene the three-step test. Moreover, other aspects of the three-step test – including that no unreasonable prejudice is caused to legitimate interests of authors by allowing AC development – needs to be satisfied. On the other conditions, *see e.g.* Senftleben (2004).

[66] The requirement in the DSM Directive that TDM should not be technically prevented only applies to non-commercial TDM activities (Art. 3(3) DSM Directive). Moreover, provisions that render void or ineffective contractual terms that limit the beneficiaries of a copyright exception do not apply to AC development permitted by the TDM exception in Art. 4 DSM Directive (*see* Art. 7 DSM Directive) or the

EU antitrust law can address these types of obstacles to AC development – by requiring copyright holders to provide access to works and grant licences for their use – and thus complements the copyright exceptions in important ways. However, antitrust can only enable and facilitate access to and use of works when the practices concerned constitute abuse of a dominant position or a restrictive agreement, which is not automatically or even generally the case when access to or use of copyright-protected materials is limited. The circumstances in which antitrust could facilitate access to and use of copyright-protected training materials in AC development, and the limitations of antitrust in this regard, are examined below.

4.1 Refusal to Grant a Licence or Access to Data for Artificial Creativity Purposes

The most direct way in which antitrust can provide access to and permit use of copyright-protected training materials is by prohibiting firms from refusing such access/use. According to CJEU case-law, an IP holder dominant in a product market may abuse a dominant position when 1) its refusal to allow access/use prevents a licensee from offering a new product not offered by the IP holder, 2) the IP in question is indispensable for operating in a downstream market, and 3) its refusal eliminates all (effective) competition in the market. The IP holder can objectively justify its conduct on grounds such as that refusal is necessary – for instance, to maintain the dominant undertaking's incentives to innovate.[67]

These criteria for abuse could be met if a copyright holder dominant in a product market refused to grant an AC developer a licence or provide access to materials in its control. Refusal to grant licences for development of an AC application can undoubtedly prevent the emergence of a new product, such as new content,

---

Footnote 66 continued

temporary reproduction exemption in Art. 5 InfoSoc Directive. Nor do the obligations of Member States to ensure that beneficiaries of certain exceptions are able to enjoy those exceptions, where hindered by technological measures, apply to temporary copying (*see* Art. 6(4) InfoSoc Directive). However, the obligations do seem to apply to the TDM exception (Art. 6(4) InfoSoc Directive in conjunction with Art. 7(2) DSM Directive), meaning that rightholders could be required to allow use of works under the TDM exception. However, this is rendered more or less meaningless by the fact that rightholders can disallow the exception from applying by making an express reservation to that effect. Moreover, as noted by Margoni and Kretschmer 2022, 695–696, the mechanism in Art. 6(4) InfoSoc Directive does not apply either when a rightholder makes works available to the public.

[67] *See e.g.* C-418/01 *IMS Health v. NDC Health* EU:C:2004:25; C-241/91 and C-242/91 *Radio Telefis Eireann (RTE) and Independent Television Publications Ltd (ITP) v. Commission of the European Communities* EU:C:1995:98; T-201/04 *Microsoft Corp v. Commission* EU:T:2007:289.

applications or services.[68] The requirement of a new product being denied to the detriment of consumers or technical development can thus frequently be met.[69]

However, the other requirements for abusive refusals are less likely to be satisfied in the context of artificial creativity applications. First, few copyright holders are dominant in any market and not all are even present in a product market (e.g. a specific content market) downstream of the rights, but simply license their works to others. Without an IP holder's presence in a product market downstream of the IP, abuse under this line of case-law is not possible since it is limited to refusals that foreclose the IP holder's rivals in a market downstream of the actual or hypothetical IP market.[70]

Second, a licence is not necessarily indispensable for AC developers to operate in the product market.[71] While substantial bodies of copyright-protected data may be needed to train artificial creativity applications, suitable training materials may be available from other copyright holders or could sometimes be produced by the AC developer itself. Operating in a product market (e.g. news articles) may also be possible by producing content by conventional means, without AC technologies and a need for licences, unless the product market is limited to AC-created content or AC-based services or applications.

Third, it is unusual for a copyright holder to be capable of eliminating all (effective) competition in the product market by refusing to grant licences. In particular, AC-based content and services would often likely compete with conventionally produced content and services that carry it; if that is the case, competition may remain intense even when content produced using AC cannot be offered on the market. However, elimination of competition would appear possible when the product market that the AC developer is seeking to operate on is specific to AC-generated content or AC-based services or applications (e.g. personalised content created automatically). Generally, the risk of eliminating competition is greater the higher the dominant firm's market share, and elimination of competition is even likely when the other criteria mentioned above are met.[72] It is not ruled out,

---

[68] C-418/01 *IMS Health v. NDC Health* EU:C:2004:25, para. 49; T-201/04 *Microsoft Corp v. Commission* EU:T:2007:289, para. 658. *See* on applicability of the case-law to data *e.g.* Graef (2016), Ch. 7; French and German Competition Authorities (2016); Crémer et al. (2019); Drexl (2017).

[69] An exception could be situations where the AC application prevented by a refusal would merely produce close copies of the works used in its development, as those may fail to serve unmet consumer demand and may in the long run undermine creation in the market. C-418/01 *IMS Health v. NDC Health* EU:C:2004:25, para. 49; T-201/04 *Microsoft Corp v. Commission* EU:T:2007:289, para. 658. *See also* Guidance on the Commission's enforcement priorities in applying Art. 82 EC Treaty to abusive exclusionary conduct by dominant undertakings, [2009] OJ EU C45/7 (24.2.2009) [Art. 82 Guidance Paper], paras. 86–87.

[70] C-418/01 *IMS Health v. NDC Health* EU:C:2004:2 (requiring at least a hypothetical upstream IP market).

[71] C-418/01 *IMS Health v. NDC Health* EU:C:2004:2; C-7/97 *Bronner* EU:C:1998:569; C-165/19 P *Slovak Telekom* EU:C:2021:239.

[72] Guidance paper, para. 85. Graef (2016), p. 273, considers that the case-law on refusals requires that dominance and exclusion take place on the same market.

though, that abuse could be found in another product market (e.g. AC-based services) in which the IP holder is not (yet) dominant owing to the risk of eliminating competition on that market.[73] Fourth, a prima facie abusive refusal could be objectively justified if the AC application concerned were to undermine incentives to create by producing outputs that infringe or freeride on the training data.[74]

Accordingly, EU antitrust can secure access to licences and supply of associated materials for AC developers where a copyright holder is dominant in a product market and is in a position to eliminate competition in the market where the AC developer operates or intends to operate. This goes beyond the prerogatives of copyright exceptions in that antitrust may enable a broader range of AC development and exploitation activities (e.g. not deemed as TDM): it can also address technical and contractual restraints on access and use as well as apply to any kind of subject matter. Importantly, antitrust can enable access even if rightholders make a reservation preventing the TDM exception from applying. However, these antitrust duties only exceptionally apply to artificial creativity pursuits, for the reasons noted above, and particularly because few copyright holders are dominant and capable of eliminating competition on a related product market.[75]

4.2 Restraints on Development or Use of AC Applications

The prohibition on abusive refusals to license is not the only means by which EU antitrust can facilitate access to copyright-protected materials for AC developers. In particular, antitrust can facilitate their access to training materials by prohibiting abusive or restrictive licensing conditions, as explained below.

4.2.1 Abusively Unfair Pricing or Conditions

Imposing excessive licensing fees or unreasonable limitations on AC development may constitute abuse of a dominant position in the IP licensing market.[76] Pricing may be abusively excessive when the prices demanded depart significantly from the economic value of the product[77] or pricing is based on a model that goes beyond what is necessary for addressing the legitimate interests of the dominant firm.[78]

---

[73] See e.g. Microsoft (2007), paras. 561–564 (risk of eliminating competition on a related market).

[74] In particular, that may amount to "duplicating" the goods and services concerned. T-201/04 *Microsoft Corp v. Commission* EU:T:2007:289, paras. 659 and 700–701. Moreover, such practices could be concluded to be harmful to innovation and consumers in the long run, thus favouring against intervention or a finding of abuse. See Art. 82 Guidance Paper, paras. 86–90.

[75] There is ongoing policy and scholarly debate on whether access to data should be mandated beyond situations governed by this line of case-law. See e.g. Crémer et al. (2019) and Drexl (2017).

[76] On the possibilities of abusive data licensing practices, see also the French and German Competition Authorities (2016).

[77] C-177/16 *Autortiesību un komunicēšanās konsultāciju aģentūra/Latvijas Autoru apvienība v. Konkurences padome* EU:C:2017:689.

[78] C-52/07 *Kanal 5 Ltd and TV 4 AB v. Föreningen Svenska Tonsättares Internationella Musikbyrå (STIM)* EU:C:2008:703.

Licensing fees demanded by dominant copyright holders could thus be abusive when pricing bears an insufficient relationship to the value of the licensed works in the context of AC applications or does not take appropriate account of the use of works. In particular, licensing fees that capture the value of the data overall, instead of just the copyright aspects, could give rise to allegations of excessive pricing, as there may be a significant discrepancy between the value of these two assets (access to data versus copyright licence).[79]

Abusively unreasonable conditions and terms (unrelated to price) may be involved if the terms go beyond serving the legitimate interest of the dominant firm and thus disproportionally limit the other party's ability to operate in the market.[80] For instance, limiting the use of works in AC development or the uses of developed AC applications could constitute such an unreasonable condition when it goes beyond what is necessary to protect the legitimate interests of copyright holders and thus unnecessarily limits the possibilities of licensees to develop new technologies and products.[81]

Notably, unlike abusive refusals to license, these types of abuse require no presence, dominance, or exclusion in the downstream product market. For this reason, abuses could more frequently apply in the artificial creativity context, as dominance in an IP market suffices. For example, copyright collecting societies or others dominant in a licensing market could be required to grant licences for artificial creativity application purposes under terms that do not constitute unfairly

---

[79] Whether pricing bears an insufficient relationship to the economic value of data raises interesting questions about the extent to which rightholders can claim compensation for the value that the data possesses but that is not attributable to its copyright-protected aspects. By virtue of their exclusive rights, copyright holders may in practice appropriate the value of data that lies in its non-protected aspects, as these non-protected aspects often cannot be exploited without infringing protected aspects. For example, copyright-protected texts may be valuable for producing grammatically appropriate text, but this aspect (grammar rules) is not protected by copyright and might be the result of efforts by parties other than the copyright holders (*e.g.* editors of the text). See e.g. C-385/07 *Der Grüne Punkt - Duales System Deutschland GmbH v. Commission* EU:C:2009:456 (unfair to demand payment for services not received). However, copyright may also allow value from such efforts not protected under copyright to be captured (*e.g.* publishers funding editing) and thus safeguard incentives relating to them. Moreover, licensing fees may be justified even where no intellectual property right is infringed, *e.g.* owing to the legal certainty a licence creates during product development. C-567/14 *Genentech Inc. v. Hoechst GmbH and Sanofi-Aventis Deutschland GmbH* EU:C:2016:526.

[80] Case C-127/73 *Belgische Radio en Televisie and société belge des auteurs, compositeurs et éditeurs v. SV SABAM and NV Fonior* EU:C:1974:25.

[81] In establishing such exploitative abuse, a central factor is whether the conditions or limitations imposed on the other party protect any legitimate interest of the dominant firm and whether less restrictive alternatives would suffice to protect them. Case C-127/73 *Belgische Radio en Televisie and société belge des auteurs, compositeurs et éditeurs v. SV SABAM and NV Fonior* EU:C:1974:25; C-52/07 *Kanal 5 Ltd and TV 4 AB v. Föreningen Svenska Tonsättares Internationella Musikbyrå* (STIM) EU:C:2008:703.

[82] Collecting societies may be required to grant licences under other rules. Although EU antitrust law does not necessarily directly result in such a duty to license, Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market, OJ EU L84/72 (20.3.2014), sets requirements on collecting societies as regards their dealings with users of works, including a duty to negotiate in good faith and to give reasons when a licence is not granted (Art. 16). Some national laws also require collecting societies to grant licences. *See e.g.* German Act on the Management of Copyright and Related Rights by Collecting Societies (17 July 2017), Sec. 34.

high licensing fees or unreasonable restraints on AC development.[82] Since proportionally protecting the interests and incentives of copyright holders may preclude a finding of abuse or provide an objective justification, rightholders may still safeguard themselves with reasonable licensing conditions against the development and use of AC applications that would threaten their interests and incentives.[83]

4.2.2 Agreements Restricting AC Development or Use

Another way in which antitrust could facilitate AC development is by prohibiting and rendering void contractual restrictions on AC development or use. For instance, this would enable AC development when the developer has access to certain material (e.g. available on the web) but cannot use it in AC development due to contractual restrictions preventing this (e.g. terms of use). Since this applies to agreements between any undertakings, not only those imposed by dominant copyright holders in the form of the abuses discussed above, this prohibition applies to a broader set of undertakings than those examined above.

Certain types of agreement that limit the use of works in AC development or their subsequent use amount to unlawful restrictions of competition. First, licensing terms that prevent the use of licensed content for AC development purposes can constitute restraints on the ability of the licensee to carry out research and development (R&D). These kinds of restraint are deemed problematic in the context of technology transfer agreements, where they amount to hardcore restrictions of competition when entered into between competitors, while agreements between non-competitors need to be assessed on a case-by-case basis,[84] regardless of whether the restricted R&D is covered by the licence or not.[85]

Copyright licensing agreements that impose restrictions on AC development may restrict competition in a manner analogous to R&D restrictions.[86] Nonetheless, the

---

[83] On assessment of abuse, *see e.g.* Case C-127/73 *Belgische Radio en Televisie et société belge des auteurs, compositeurs et éditeurs v. SV SABAM and NV Fonior* EU:C:1974:25; C-52/07 *Kanal 5 Ltd and TV 4 AB v. Föreningen Svenska Tonsättares Internationella Musikbyrå (STIM)* EU:C:2008:703. Moreover, as with refusals to license, in cases of prima facie abuse, rightholders could be able to justify pricing and other unfair conditions in the context of artificial creativity applications on the basis that such conditions are necessary for protecting their own incentives to create content. For example, the risk of an AC application harming the expected rewards of the copyright holder could potentially justify imposing conditions on – or higher licensing fees for – those artificial creativity applications that produce outputs containing protected aspects or that otherwise harm incentives to create. On efficiency and other justifications for abuse, *see* C-209/10 *Post Danmark A/S v. Konkurrencerådet* EU:C:2012:172; Guidance Paper, pp. 28–31.

[84] Commission Regulation (EU) No 316/2014 of 21 March 2014 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of technology transfer agreements, OJ EU L93/17 (28.3.2014) (TTBER), Arts. 4(1)(d) and 5(2).

[85] Commission, Guidelines on the application of Art. 101 of the Treaty on the Functioning of the European Union to technology transfer agreements, OJ EU C89/3 (28.3.2014) (Guidelines), para. 115.

[86] Although the TTBER does not apply to the licensing of copyright-protected content for AC development (as TTBER only applies to licensing of technology for production purposes), the principles applicable to technology transfer may provide guidance. Copyright licensed in an AC context can be seen as of a technical nature since it is only used functionally, to train neural networks.

mere fact that a licence does not permit the use of works in AC development does not render it comparable to such an R&D restriction. For instance, if a licence has been granted only for consuming works in certain ways (e.g. non-commercially), it does not follow that the agreement restricts R&D – it merely does not authorise R&D in the first place.[87] Second, restraints on the ability to use AC applications developed can also restrict competition in the product market (e.g. for content or services). For example, restrictions on offering AC-generated content in certain geographical areas could even constitute a presumptively banned passive territorial sales restraint.[88]

Accordingly, antitrust liberates licensees from restrictions on the development of AC applications or their subsequent use where such restraints restrict competition in the above manner. However, non-dominant rightholders not willing to authorise use of their works in AC development can simply refuse to enter into licensing agreements covering AC development, while dominant ones can generally also do so, except in the exceptional circumstances discussed above, without explicitly needing to restrict AC development.[89] Therefore, these antitrust prohibitions on agreements that restrict AC development or use would mostly be relevant if AC developers were already parties to a licensing agreement that would allow AC development (were it not for an explicit restriction on doing so) or could enter into such a licensing agreement.

5 AC Development under EU Copyright and Antitrust: Gaps in the Current Framework and Impact of EU Legislative Initiatives Relating to Data Access

AC development is possible, enabled and facilitated by EU copyright and antitrust law in the ways discussed above. Individual access and use enablement and facilitation mechanisms in the laws complement each other to some extent but, as noted below, the overall framework is not capable of effectively addressing various obstacles to AC development – even where AC development is in principle possible under the copyright exceptions examined above. As new and upcoming EU legislative initiatives dealing with access to data also only marginally apply to the types of training data that AC development requires, this begets the question whether unaddressed obstacles to AC development warrant policy intervention, as discussed next.

---

[87] Whether copyright exceptions could authorise uses beyond the licences granted may not be clear. For example, the TDM exception might be inapplicable if TDM is expressly reserved by rightholders or the use contravenes the requirement that the works or other subject matter be lawfully accessible, where access is not permitted for TDM purposes. *See* Art. 4(1) DSM Directive.

[88] In particular if the outputs do not infringe the works used in AC development, then even copyright itself would not prevent sales. For a comparison, *see e.g.* C-429/08 *Football Association Premier League and Others*, EU:C:2011:63.

[89] Moreover, restrictions could also be justified particularly when they are reasonably necessary, e.g. to protect the incentives of copyright holders to create under Art. 101(3) TFEU. *See e.g.* C-501/06 P, C-513/06 P, C-515/06 P and C-519/06 P *GlaxoSmithKline Services Unlimited v. Commission of the European Communities* EU:C:2009:610.

5.1 Boundaries of Copyright and Antitrust Mechanisms

While copyright and antitrust facilitate access to and the ability to use training data in AC development, their capability to effectively do so is limited to particular types of AC applications and certain types of training materials. In addition, they are subject to various technical requirements and other conditions and, moreover, face legal uncertainty in many respects, as observed above. In particular, while the TDM exception does quite broadly cover AC development, the effective possibilities thereof can be curtailed by rightholders reserving the mining right as well as by the uncertainty surrounding whether legally recognised reservations have been made and whether the materials to be used were lawfully accessible. While the temporary copying exemption complements the TDM exception by providing an additional legal basis for AC development and omitting the possibility of rightholder reservations, it does not cover all categories of subject matter and imposes certain requirements on the process of AC development (e.g. on expunging used and resulting copies) in ways that limit its applicability for AC development. Furthermore, while the three-step test does not pose an obstacle where the outputs of the AC developed would not infringe or significantly freeride on the training materials, other kinds of AC applications may fail the three-step test.[90]

What is more, although EU antitrust law entitles AC developers to gain access and to use copyright-protected training materials, this only applies in rare circumstances where a copyright holder is dominant and able to reserve another market to itself by excluding the AC developer from it. While access and use are facilitated by prohibitions on restraints and unfair pricing that limit AC development or use, these bans are chiefly effective when a licensor is dominant on a licensing market (e.g. as copyright collecting societies often are).

The legally facilitated or safeguarded possibilities, entitlements and rights to access and use protected subject matter for the purposes of AC development further depend on what obstacles AC developers seek to overcome: copyright infringement or violations, contractual restraints or technical obstacles. Notably, both copyright and antitrust are regularly powerless against contractual and technical restraints that prevent access to or use of works for AC development. EU copyright law does not address technical measures or other obstacles or contractual restraints that prevent the beneficiaries of the temporary copying exemption or TDM exception (Art. 4 DSM Directive) from accessing or using works for the exempted purposes.[91] While Art. 7(2) DSM Directive in conjunction with Art. 6(4) InfoSoc Directive obliges

---

[90] In these respects, Ørstavik I (2022) and Senftleben (2022) are more concerned than this article is about the scope and applicability of the TDM and temporary copying exceptions. The argument made in this article is that most of the substantive copyright issues and related uncertainties can be largely resolved through interpretations that are fully consistent with existing CJEU case-law and the language of the Directives.

[91] The DSM Directive only provides for contractual ineffectiveness (Art. 7(1) DSM Directive) and that access should not be unreasonably limited where TDM carried out for scientific purposes is concerned (Art. 3(3) DSM Directive). Although some aspects of AC development could be carried out by qualifying research organisations for scientific purposes, the strict conditions of this exception do not allow for commercially driven AC development, and further commercial development and the use of neural networks developed under the scientific TDM exception can also be problematic, as noted in *supra* note 13.

Member States to enable use of works for the purpose of TDM under Art. 4 DSM Directive if prevented by technical measures, this has little practical significance, as rightholders can preclude the TDM exception from applying by making an express reservation. Moreover, the first subparagraph of Art. 6(4) InfoSoc Directive only concerns technological measures, and thus does not oblige Member States to take measures to enable access and use where the obstacles faced are of some other technical practical nature (e.g. inability to gain access to content at all). Additionally, EU antitrust law is in these respects toothless against unilateral practices by non-dominant rightholders, many agreements limiting use of works in AC development, and most practices by dominant firms, as noted above.[92]

Accordingly, the EU copyright and antitrust framework lacks tools to address technical and contractual obstacles to the use of works in AC development, including when this would be allowed under the copyright exceptions discussed above. The more content is only available in an access-controlled, technically protected manner or subject to contractual restraints, the more these irremediable obstacles to gaining access to training materials hamper the ability to carry out AC development that would be permitted in copyright law.

5.2 Implications of EU Initiatives to Improve Data Access

The EU has recently adopted and proposed pieces of legislation that grant access to data in certain situations. However, these only apply in limited scope to training materials that could be used to develop AC applications or create an access mechanism that benefits AC developers. First, the proposed Data Act is limited to the user's own (raw) data and applies only to products and services that rarely generate creative content.[93] However, it is relevant for some types of user-generated content that can be used as training data: the Data Act would enable third-party services to gain access to data at the user's request.[94] Moreover, the Data Act renders unfair contracts related to data access unenforceable, which may liberate AC developers from unreasonable restraints.[95]

Second, the Data Governance Act does not create new data access obligations that benefit AC developers directly but facilitates access in other ways by creating infrastructures and preventing restraints on access to public data.[96] Nonetheless, it

---

[92] It is only in extreme circumstances that antitrust obliges copyright holders to provide access to their works for the purposes of AC development.

[93] 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/2022/68 final. In the initial Commission proposal, the scope of products covered was broader so as to cover all kinds of connected devices, including ones that generate audiovisual or other content. A more recent draft would exclude connected products whose primary purpose is to play, record or transmit content. *See e.g.* First Presidency compromise text, Council document 11194/22 (12 July 2022).

[94] *E.g.* this could apply to videos, photos or audio recordings taken by users.

[95] Article 13 Data Act.

[96] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724, OJ EU L 152/1 (3.6.2022) (Data Governance Act).

can safeguard and facilitate access to certain data held by public-sector bodies[97] that could be used to generate some types of content, by banning exclusivity arrangements and unreasonable restrictions on reusing the data as well as the unreasonable pricing of data.[98] For instance, public-sector bodies may possess written text in various forms and images that could be relevant for some AC development purposes.

The Digital Markets Act does not create any general obligations to grant access to data even to core platform services. However, it does require platforms to enable user data portability and to grant access to business users of platforms to certain data generated in the latter's services.[99] This could give AC developers access to data provided by or generated by platform users (e.g. photos or other content uploaded by users) when users exercise their portability right.

Accordingly, the Acts proposed or adopted facilitate access only to certain types of (mostly user-provided or user-generated) data, which may include forms of content such as photos, video or sound recordings, or written text. This is because of the limited scope of application of the Acts and their data-specific obligations, which leave much creative content outside their access mechanisms (other than user-generated content). The Acts do not directly authorise AC developers to access such data but allow them to access user data only if users exercise their access and portability rights to give AC developers such access. Moreover, the Acts do not authorise third parties to gain access to user data to carry out copyright-relevant acts. AC developers would need to rely, for instance, on a copyright exception or licence from the copyright holders in order to be able to carry out AC development employing the user data. Overall, these EU legislative initiatives grant AC developers only very limited access to content that could be used in AC development.

5.3 Need for Intervention?

These gaps in substantive scope and effectiveness raise the question whether a case is to be made for securing access to copyright-protected data through mechanisms similar to those that apply in copyright law to technological protection measures, access/use restrictions, and agreements that prevent beneficiaries of certain other copyright exceptions from being able to carry out exempted activities, or by adjusting the applicable antitrust standards in this regard, or by other means (e.g. data-specific regulation). In particular, the need for such policy interventions depends on the extent to which AC development is possible under the current framework and how effectively voluntary licensing enables AC development. In particular, licensing models already exist (e.g. Creative Commons) and governance

---

[97] Yet, *e.g.* public broadcasters and cultural and education bodies are excluded from the scope of certain main rules. Article 3(2) Data Governance Act.

[98] Articles 4–6 Data Governance Act.

[99] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828, OJ EU L 265/1 (12.10.2022) (Digital Markets Act), Art. 6(9) and (10). New obligations relating to data can also be created by the Commission under Art. 12 of the Act.

mechanisms for controlling the use of works in AC development are emerging that could address copyright infringement issues, while bilateral agreements between the holders of rights to large bodies of subject matter and AC developers can also enable AC development.

However, it remains a concern that the current regime might limit the ability to engage in AC development to organisations that already have access to and rights in respect of portfolios of content (e.g. licensed content or user-provided content), those capable of taking the risk of potentially major copyright infringement and the associated costs of litigation, and those carrying it out outside of the EU.[100] Such an outcome would be detrimental to technical development and content production. Nor would it necessarily even serve the interests of those creating creative content if their works are used in any case (e.g. outside of the EU, without meeting the conditions that the copyright exceptions examined above impose on lawful access or rightholder reservations) and as the lack of rivalry in AC development may further strengthen the bargaining and market positions of the organisations and other parties that can currently engage in AC development.

6 Conclusions

Use of AI to produce content has significant potential to facilitate and automate production of content as well as to offer entirely new kinds of content and content services. However, risks of copyright infringement may hamper the development of such technologies and applications. While the exception for TDM can apply to their development, the possibility of rightholder reservations prevents the exception applying. At the same time, ensuring that no such reservation has been made, as well as that the materials were lawfully obtained, is virtually impossible unless these two conditions are interpreted in ways that support the emergence of practical solutions to these issues. Although the exemption on temporary copying applies regardless of rightholder reservations, it does not cover all relevant categories of protected subject matter and entails other limitations and uncertainties that limit its applicability in AC development. Both exceptions can also only be applied by national courts if they satisfy the three-step test, which may not be the case if the AC applications were to produce infringing or freeriding outputs.

Even though copyright within the above boundaries would allow AC applications to be developed, copyright, antitrust and other recently adopted or proposed EU legislation are largely incapable of addressing the technical and contractual obstacles that AC developers face. Antitrust confers access to copyright-protected materials only in rare situations – mostly where a dominant rightholder is concerned – and EU legislative initiatives mostly concern user-provided or user-generated content to which AC developers can gain access only if users exercise their data portability rights. These obstacles – still unaddressed in the current and proposed

---

[100] Use of works to develop AI-based content production can constitute fair use under U.S. copyright, depending on a case-by-case assessment. Lemley and Casey (2021). In the United Kingdom, the government supports the introduction of a TDM exception that offers more leeway than the EU one, including not giving rightholders the possibility of reserving mining. UKIPO (2022).

legal framework – risk hampering AC development, to the detriment of technical development and content production, by allowing only certain organisations to develop them. The existing framework does not safeguard incentives to create or serve rightholder interests either, as the regime may shift AC development outside of the EU, where it need not be carried out in compliance with the conditions set by EU copyright exceptions and where the conditions do not incentivise the development of emerging governance models that allow rightholders to control the use of their subject matter in AC development.

References

Aggarwal C (2015) Data mining the textbook. https://doi.org/10.1007/978-3-319-14142-8

Brown TB et al. (2020) Language models are few-shot learners. arXiv:2005.14165 [cs.CL]. Accessed 30 Nov 2022

Chiou T (2020) Copyright lessons on machine learning: what impact on algorithmic art? 10 (2020) JIPITEC 398 para. 1. https://nbn-resolving.org/urn:nbn:de:0009-29-50250. Accessed 30 Nov 2022

Crémer J et al. (2019) Competition policy for the digital era. https://doi.org/10.2763/407537

Drexl J (2017) Designing competitive markets for industrial data—between propertisation and access. JIPITEC 8:257 para. 1 https://nbn-resolving.org/urn:nbn:de:0009-29-46365. Accessed 30 Nov 2022

Drexl J et al (2019) Technical aspects of artificial intelligence: an understanding from an intellectual property law perspective (8 October 2019). Max Planck Institute for Innovation and Competition Research Paper No. 19-13 https://doi.org/10.2139/ssrn.3465577

Drexl J et al. (2021) Artificial intelligence and intellectual property law - position statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the current debate (9 April 2021). Max Planck Institute for Innovation and Competition Research Paper No. 21-10 https://doi.org/10.2139/ssrn.3822924

Ducato R, Strowel A (2021) Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. EIPR 43(5):322–337

European Parliament (2018) The exception for text and data mining (TDM) in the proposed Directive on Copyright in the Digital Single Market – legal aspects http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf. Accessed 30 Nov 2022

Franceschelli G, Musolesi M (2021) Copyright in generative deep learning. arXiv:2105.09266 [cs.CY]. Accessed 30 Nov 2022

French Competition Authority and the German Federal Cartel Office (2016), Competition law and data https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2. Accessed 30 Nov 2022

Frosio G, Bulayenko O (2018) The exception for text and data mining (TDM) in the proposed Directive on Copyright in the Digital Single Market—Legal Aspects. (February 2018)

Gervais D (2019) Exploring the interfaces between big data and intellectual property law. 10 (2019) JIPITEC 3 para. 1 https://nbn-resolving.org/urn:nbn:de:0009-29-48753. Accessed 30 Nov 2022

Geiger C, Gervais D, Senftleben M (2014) The three-step test revisited: how to use the test's flexibility in national copyright law. Am Univ Int L Rev 29(3):581–626. https://doi.org/10.2139/ssrn.2356619

Geiger C, Frosio G, Bulayenko O (2018a) The EU Commission's proposal to reform copyright limitations: a good but far too timid step in the right direction. EIPR 40(1):4–15

Geiger C, Frosio G, Bulayenko O (2018b) The exception for text and data mining (TDM) in the proposed Directive on Copyright in the Digital Single Market—Legal Aspects https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf. Accessed 30 Nov 2022

Gillotte J (2020) Copyright infringement in AI-generated artworks. (2020) 53 UC Davis L Rev 2655

Gordon W (2003) Excuse and justification in the law of fair use: transaction costs have always been part of the story. 50 J. Copyright Soc'y USA 149 (2002–2003)

Graef I (2016) EU competition law, data protection and online platforms: data as essential facility. Wolters Kluwer (ISBN 9789041183248)

Griffiths J, Synodinou T, Xalabarder R (2023) Comment of the European Copyright Society addressing selected aspects of the implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market. GRUR International 72(1):22–36. https://doi.org/10.1093/grurint/ikac137

Grimmelmann J (2016) Copyright for literate robots. 101 Iowa L Rev 657 (2016)

Guibault, L, Westkamp, G, Rieber-Mohn, T (2007) Study on the implementation and effect in member states' laws of Directive 2001/29/EC on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society. (IVIR).

Hartmann C et al. (2021) Trends and developments in artificial intelligence communications networks, content and technology final report prepared by: challenges to the intellectual property rights framework. https://doi.org/10.2759/458120

Hilty R, Richter H (2017) Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3 – Text and Data Mining) (14 January 2017). Max Planck Institute for Innovation and Competition Research Paper No. 17-02 https://doi.org/10.2139/ssrn.2900110

Iglesias PM et al. (2021) Intellectual property and artificial intelligence—a literature review. EUR 30017 EN, Publications Office of the European Union, Luxembourg

Jackson MB (2021) Text and data mining exception in South America: a way to foster AI development in the region. GRUR International 70(12):1145–1157. https://doi.org/10.1093/grurint/ikab081

Kop M (2020) Machine learning & EU data sharing practices (3 March 2020). Stanford—Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford Univ, Issue No. 1/2020 https://ssrn.com/abstract=3409712. Accessed 30 Nov 2022

Lemley M, Casey B (2021) Fair learning. 99 Texas L Rev 743

Levendowski A (2018) How copyright law can fix artificial intelligence's implicit bias problem. 93 Wash L Rev 579

Margoni T (2018) Artificial intelligence, machine learning and EU copyright law: who owns AI? (10 November 2018). CREATe Working Paper 2018/12 (December 2018) https://doi.org/10.2139/ssrn.3299523

Margoni T, Kretschmer T (2022) A DEEPER LOOK INTO the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. GRUR International 71(8):685–701. https://doi.org/10.1093/grurint/ikac054

Meys R (2021) Data mining under the Directive on Copyright and Related Rights in the Digital Single Market: are European database protection rules still threatening the development of artificial intelligence? GRUR International 69(5):457–473. https://doi.org/10.1093/grurint/ikaa046

Ohly A (2010) The freedom of imitation and its limits—a European perspective. Int Rev Intellect Prop Compet Law 41:506–524

Ørstavik I (2022) Access to data for training algorithms in machine learning: copyright law and right-stacking. In: Pihlajarinne T, Alen-Savikko A (eds) Artificial intelligence and the media—reconsidering rights and responsibilities. Edward Elgar. https://doi.org/10.4337/9781839109973

Proposal for adirective of the European parliament and of the council on copyright in the Digital Single Market. COM/2016/0593final

Rosati E (2019) Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity (12 September 2019). Asia Pac L Rev. https://doi.org/10.1080/10192557.2019.1705525

Sag M (2009) Copyright and copy-reliant technology (9 April 2009). Northwestern Univ L Rev, Vol. 103

Sartor G, Lagioia F, Contissa G (2018) The use of copyrighted works by AI Systems: Art works in the data mill (October 11, 2018) https://doi.org/10.2139/ssrn.3264742

Scheuerer S (2021) Artificial intelligence and unfair competition—unveiling an underestimated building block of the AI regulation landscape. GRUR International 70(9):834–845. https://doi.org/10.1093/grurint/ikab021

Schönberger D (2018) Deep copyright: up – and downstream questions related to artificial intelligence (AI) and machine learning (ML) (9 January 2018). In: De Werra J (ed) Droit d'auteur 4.0 / Copyright 4.0. Geneva / Zurich (Schulthess Editions Romandes) 2018, 145–173 https://archive-ouverte.unige.ch/unige:101505. Accessed 30 Nov 2022

Senftleben M (2004) Copyright, limitations and the three-step test. An analysis of the three-step test in international and EC copyright law. Kluwer Law International

Senftleben M (2022) Compliance of national TDM rules with international copyright law: an overrated nonissue? IIC 53:1477–1505. https://doi.org/10.1007/s40319-022-01266-8

Sobel B (2017) Artificial intelligence's fair use crisis. 41 Columbia J of Law & the Arts. 45 https://doi.org/10.7916/jla.v41i1.2036

Sobel B (2021) A taxonomy of training data: disentangling the mismatched rights, remedies, and rationales for restricting machine learning. In: Hilty RM, Lee J-A, Liu K-C (eds) Artificial intelligence and intellectual property. Oxford University Press. https://doi.org/10.1093/oso/9780198870944.003.0011

Synodinou T (2019) Lawfulness for users in European copyright law: acquis and perspectives. 10 JIPITEC 20 para. 1 urn:nbn:de:0009-29-48767

Triaille et al. (2014) Study on the legal framework of text and data mining (TDM). https://doi.org/10.2780/1475

UKIPO (2022) Artificial intelligence and intellectual property: copyright and patents: Government response to consultation (28 June 2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation. Accessed 30 Nov 2022

Vesala J (2022) Press publishers' right and artificial intelligence (14 January 2021). In: Pihlajarinne T, Alen-Savikko A (eds) Artificial intelligence and the media – reconsidering rights and responsibilities. Edward Elgar. https://doi.org/10.4337/9781839109973

Vesala J, Ballardini RM (2019) AI and IPR infringement: a case study on training and using neural networks. In: Ballardini RM, Pitkänen O, Kuoppamäki P (eds) Regulating industrial internet through IPR Data Protection and Competition Law. Kluwer Law International, pp 99–114