The Challenge of Balancing Artistic Autonomy and AI Training

– Evaluating the Effectiveness of the Opt-Out Mechanism under Art 4(3) DSM Directive for Artist Protection

Anna Buss

ABSTRACT

The introduction of Article 4 of the DSM Directive was intended to create a commercial exception for text and data mining. The intention behind the article was to foster innovation and to create a legal framework to accommodate this. However, the article is not very well drafted, particularly in relation to the rise of generative AI and the training of AI systems. The article contains a reservation of use clause that allows authorsauthors to opt-out of the use of their works for text and data mining. The absence of an EU standard of opt-out declaration creates certain complications.

The AI Act initially seemed to constitute a promising solution, but the opportunity to eliminate legal uncertainties was not utilised. Instead, there is a high chance of blocking future technological innovations.

1. INTRODUCTION

Since the launch of ChatGPT by OpenAI in November 2022, legal issues in the field of AI have become more relevant than ever.

The special feature of generative AI is that the outputs can hardly be distinguished from human creations. For instance, a canvas print in the style of the Old Masters[1] created by AI was sold for almost half a million dollars at a Christie's auction in 2018.[2]

However, in order to understand how these outputs are produced, one needs to looked into how the AI systems work on the input side.

The interest in these matters is growing rapidly in the legal literature and research. Therefore, there is a great need for legal clarity.

While, authors are calling for bans, remuneration and transparency,[3] AI developers are pointing out that overly strict regulatory requirements could make the EU increasingly unattractive as a business location for development of AI technology.[4]

In order to evaluate the legal framework that is to regulate AI, it is of importance to look into the technical perspectives and understand what is covered by the term AI,

---

[1] The term "Old Master" is used to describe famous European authors who existed between about 1300 and 1800, covering the Early Renaissance to the Romantic period via "Old Masters" (*The Art Story*) <https://www.theartstory.org/definition/old-masters/> accessed on 04 October 2024.

[2] *Portrait of Edmond Belamy* (2018) constructed by arts-collective *Obvious*, sold for $432,500 (original estimate of $7,000-$10,000) via Allyssia Alleyne, "A sign of things to come? AI-produced artwork sells for $433K, smashing expectations" (*CNN*, 25 October 2018) <https://edition.cnn.com/style/article/obvious-ai-art-christies-auction-smart-creativity/index.html> accessed on 04 October 2024.

[3] The use of copyright protected data was called the largest art heist in history in an open letter written by a coalition of authors, journalists and actors, via "Restrict AI Illustration from Publishing: An Open Letter" (Center for Artistic Inquiry and Reporting, 2 May 2024) <https://artisticinquiry.org/AI-Open-Letter>; Other statements: Initiative Urheberrecht *(Initiative Copyright)* "Ruf nach Schutz vor generativer KI" (Initiative Urheberrecht, 19 April 2023) <https://urheber.info/diskurs/ruf-nach-schutz-vor-generativer-ki>; "Joint statement from authors' and performers' organisations on Artificial Intelligence and the AI Act" (The Federation of European Screen Directors, 9 February 2023) <https://screendirectors.eu/joint-statement-from-authors-and-performers-organisations-on-artificial-intelligence-and-the-ai-act/>; "Our Manifesto for AI companies regulation in Europe" (European Guild for Artificial Intelligence Regulation, 4 November 2023) <https://www.egair.eu/#manifesto>; Matthew Butterick and Joseph Saveri, "We've filed a lawsuit challenging Stable Diffusion, a 21 st-century collage tool that violates the rights of authors" (Image generator litigation) <https://stablediffusionlitigation.com>, all links were accessed on 04 October 2024.

[4] Enrico Bonadio, Luke McDonagh, "Artificial intelligence as producer and consumer of copyright works: evaluating the consequences of algorithmic creativity" (2020) 2 IPQ 112.

2. A TECHNICAL AND LEGAL EXAMINATION OF AI TRAINING

2.1 Technical Background

AI is an area of computer science with no universal definition.[5]

It can be said that AI is an umbrella term that encompasses various rule-based computer technologies.[6]

Machine learning describes the learning process of a computer system that teaches the system to identify new patterns in data and to apply this knowledge to new data, as well as to generate new output.[7] Broadly speaking, the learning process entails an algorithm that receives training data, reflecting past knowledge or experience, and generates information usable by other algorithms for tasks like prediction or decision-making.[8] The great capability of the system can be traced back to this artificial neural network. This form of machine learning is called "deep learning".[9] Developing machine learning presupposes the use of a large amount of data that is to be fed into the model during the training process. As an example, Stability AI used the LAION-5B data set, which consists of 5.85 billion links to filtered image-text pairs, for the training of Stable Diffusion.[10]

In order to obtain as much authentic high-value data as possible, the data is taken from the Internet by "web scraping".[11]

Due to the multi-stage process, the question of copyright infringement must be considered both in the context of the collection and storage of data for training purposes (Step 1) and in the storage of information from the data in the neural network (Step 2).

2.2 Reproduction during the Data Collecting (Step 1)

The definition of reproduction given by Article 2 of the InfoSoc Directive implies that reproduction includes any physical act capable of rendering the work directly or indirectly perceptible to the human senses.[12] Conse-quently, it is irrelevant whether this is done consciously or unconsciously.[13]

It can be concluded that the process of data collecting results in copyright-relevant action which therefore requires consent.

Most developers will not have licenses for the data used for them and therefore the use initially constitutes an infringement of the rights of the rightholders.[14]

2.3 Reproduction in the Neuronal Network (Step 2)

The question of whether storage results in reproduction is currently the subject of controversial debate.

The prevailing opinion is that the storage of information does not result in reproduction.[15] It is argued that neural networks do not contain protected works, but merely information such as patterns or correlations from the training data.

Other voices are convinced that reproduction occurs during storage.[16]

The arguments here relate to the fact that reproduction is technology-neutral. This means that if reproduction is possible, then there is *de facto* reproduction. One of the main arguments is that clever prompts can be used to get the AI to reproduce the protected works under certain circumstances.[17]

The wording of Article 2 of the InfoSoc Directive, which states: "in any way and in any form", speaks in favour of the view that storage occurs. Consequently, this must also include processes that take place in the neural network.[18]

However, this must be countered by the fact that the developers' intention is not to reproduce, but rather that the system should use the information to derive patterns and abstract relationships in order to independently create new things.[19]

22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society.

13 A subjective element is only relevant for the criminal law assessment, see Artur Wandtke, Winfried Bullinger, *Praxiskommentar Urheberrecht* (6th edn, CH Beck 2022) para. 29.

14 Jonathan Pukas, 'KI-Trainingsdaten und erweiterte kollektive Lizenzen – Generierung von Werken als KI-Trainingsdaten auf Basis erweiterter kollektiver'(2023) GRUR 614, 615.

15 Consenting: Benjamin Raue, 'Die geplanten Text und Data Mining-Schranken (§§ 44b und 60d UrhG-E)' (2020) ZUM 172, 173; Benjamin Raue, 'Rechtssicherheit für datengestützte Forschung'(2019) ZUM 684, 686; Andrea Hagemeier, *BeckOK UrhR: § 44b UrhG* (37th edn, CH Beck 2023) para. 1–3; Haimo Schack, 'Schutzgegenstand, „Ausnahmen oder Beschränkungen" des Urheberrechts' (2021) GRUR 904, 905; Niklas Maamar, 'Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen' (2023) ZUM 481, 483.

16 Dissenting: Haimo Schack, 'Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda' (2024) NJW 113, 115.

17 Malte Baumann, 'Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld'(2023) NJW 3673, 3674: Researchers have succeeded in getting AI systems to reproduce a novel word for word or to reproduce images identically.

18 Consenting: Paulina Jo Pesch, Rainer Böhme, 'Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern' (2023) GRUR 997, 999; Marcus von Welser, 'Generative KI und Urheberrechtsschranken' (2023) GRUR 516, 517.

19 Franz Hofmann, 'Retten Schranken Geschäftsmodelle generativer KI-Systeme?' (2024) ZUM 166, 167.

5   Ryan Calo, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017) 399, 404.

6   Josef Drexl, Reto Hilty et al., 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective'(2019) Max Planck Institute for Innovation and Competition Research Paper 19-13, 3 <https://ssrn.com/abstract=3465577> accessed on 04 October 2024.

7   European Commission, Artificial Intelligence for Europe, COM (2018) 237 final, 10.

8   Jyh-An Lee, Reto Hilty, Kung-Chung Liu, 'Artificial Intelligence and Intellectual Property' (2021) Oxford University Press 2021-13, 11 <https://ssrn.com/abstract=3802232> accessed on 04 October 2024.

9   Martin Kretschmer, Thomas Margoni, Pinar Oruç, 'Copyright Law and the Lifecycle of Machine Learning Models' (2024) 55 IIC 110, 114.

10 <https://laion.ai/projects/> accessed on 04 October 2024.

11 Tsaone Swaabow Thapelo et al., 'SASSCAL WebSAPI: A Web Scraping Application Programming Interface to Support Access to SASSCAL's Weather Data' (2021) 20 Data Science Journal <https://datascience.codata.org/articles/10.5334/dsj-2021-024> accessed on 04 October 2024.

12 Directive (EC) 2001/29 of the European Parliament and of the Council of

The explicit aim here is not to reproduce, but to create.[20]

A visit to a museum can inspire individuals through exposure to art pieces. If subsequent work is created based on this inspiration that meets the requirements for copyright protection, it constitutes a new copyrighted work. Consequently, there is no infringement of the artist's rights, as copyright law does not protect the right to consume a work for inspiration.[21]

This can also be applied to AI. A computer should be free to break down a work into its individual parts in order to extract information for new art from it.

Information must remain a free good as it is anchored in Article 11 of the Charter of Fundamental Rights.[22] A monopolisation of information would lead to a severe restriction of freedom of expression.

On the other hand, reproduction in the human brain cannot really be compared with the recording of information in a neural network.

AI is a tool and even if the neural network stores the information, one could come to the conclusion that this is a reproduction.

However, it must be taken into account here that Recital 9 of the DSM Directive states that pure data in the form of factual information does not constitute copyright-relevant acts.[23] It is not the protected works themselves that are stored in the trained AI model, but the information obtained from machine learning.[24] This corresponds to the free enjoyment of the work which is secured by Article 13 of the Charter of Fundamental Rights.

Overall, it must be concluded that storage in the neural network does not constitute reproduction.

2.4 Conclusion

The storage of data in neural networks does not constitute an act of reproduction within the meaning of Article 2 of the InfoSoc Directive due to the absence of copyright-relevant actions. Nevertheless, the data collection processes are clearly to be regarded as acts of reproduction.

This result is welcome. On the one hand, it establishes that information and thus the pure enjoyment of a work is freely accessible, while on the other hand, the interests of authors are taken into account. Recognising reproductions in the data collection process ensures that works may not be used without the author's permission. Therefore, the unauthorised use of the data constitutes copyright infringement.

3. THE SCOPE OF ARTICLE 4 OF THE DSM DIRECTIVE

The EU copyright system aims to create a balance between the interests of authors and the general public and users. This is done by recognising that although authors have an exclusive right to their works, their right is not unlimited. Article 3 and Article 4 of the DSM Directive contain exceptions for uses of text and data mining. Article 4 allows TDM in the case of lawfully accessible works that are not subject to a machine-readable reservation of the rights holder.

The aim of the European legislator was to remove legal uncertainty for data analysis and thus strengthen European competitiveness in (digital) markets and thus the EU as a business location.[25] Therefore, Article 4 of the DSM Directive is referred to as the commercial exception.[26]

Although AI is not explicitly mentioned in the DSM Directive, the meaning and purpose of the directive speaks in favour of it being formulated openly and consequently including AI. Any other interpretation would not be coherent and would contradict the legislator's intention, which was to clearly extend the directive to the development of new technologies.

This is supported by the wording of Recital 18, which refers to the further development of new technologies, implying thus the training of AI falls under text and data mining.

Ultimately, Article 53(1)(c) of the AI Act speaks of: "Providers of general purpose AI models shall: [...] put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790".[27]

Consequently, the training of generative AI falls under the text and data mining exception.

As a result, there is legal legitimation to collect copyright protected works during AI training and create a training corpus from them.

4. THE OPT-OUT MECHANISM UNDER ARTICLE 4(3) OF THE DSM DIRECTIVE AS AN ADEQUATE PROTECTION FOR AUTHORS

The question is whether the opt-out mechanism in Article 4(3) of the DSM is an effective protection for authors. As the provision does not specify exactly how the opt-out

20   Paulina Jo Pesch, Rainer Böhme, 'Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern' (2023) GRUR 997, 1006.

21   Jonathan Pukas, 'KI-Trainingsdaten und erweiterte kollektive Lizenzen – Generierung von Werken als KI-Trainingsdaten auf Basis erweiterter kollektiver' (2023) GRUR 614, 616.

22   Charter of Fundamental Rights of the European Union, OJ C 326, 26 October 2012.

23   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.

24   Katharina de la Durantaye, '»Garbage in, garbage out« – Die Regulierung generativer KI durch Urheberrecht' (2023) ZUM 645, 647.

25   Benjamin Raue, 'Die Freistellung von Datenanalysen durch die neuen Text und Data Mining- Schranken (§§ 44b, 60d UrhG)' (2021) ZUM 793, 794.

26   ibid.

27   Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

is to be declared, this question can only be answered by analysing the specific requirements of the provision. This is also necessary in order to identify the weaknesses of the provision in the next step.

Recital 18 of the DSM Directive indicates that the reservation by the rightholder needs to be in an appropriate manner and differentiates between content which has been made publicly available online and other cases.

With regards to content made publicly available online it is only considered appropriate by fulfilling the requirement of machine readability.[28]

The CJEU ruled in *VG Bild-Kunst* that the adoption of effective technological measures within Article 8(1) and (3) of the InfoSoc Directive is necessary.[29]

This is intended to ensure both legal certainty and the functionality of the Internet.[30]

The DSM Directive itself does not specify what machine-readability means. Therefore, this needs to be analysed with different approaches.

4.1 Machine Readability

With regard to the practicability of the opt-out mechanism, many legal scholars have already expressed doubts.[31] The question that arises is how to declare the opt-out effectively.

Article 4(3) of the DSM Directive states that the TDM exception in Article 4(1) of the DSM Directive is subject to a reservation of use by the rightholder, provided that the rightholder has declared it in an appropriate manner. As stated above, the provision refers to a declaration in machine-readable form as appropriate. The reason for this is that automated crawlers are used for data collecting.[32]

However, the term machine-readable is not defined by the directive and only little information about machine-readability is provided. Recital 18 of the DSM Directive states: "to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service", meaning that all textual forms of expression are covered.[33] Furthermore, metadata and terms and conditions of a website or a service are mentioned by way of example and not exhaustively. This is emphasised by the "including".

There is also no technical standard for machine readability within the EU yet.

Furthermore, the national legislators have not made use of their possibility to implement a definition into national legislation.[34]

It is therefore necessary to look into what constitutes "machine-readable" outside the DSM Diective

The Cambridge dictionary defines machine-readable as: "(of information or printed text) able to be understood and used by a computer."[35]

This definition is not particularly enlightening, one interpretation could be that Machine-readable could be understood to simply mean a digital expression of the opt-out. Consequently, any written language that can be digitalised would be covered.

Recital 35 of Directive 2019/1024 states:

> A document should be considered to be in a machine-readable format if it is in a file format that is structured in such a way that software applications can easily identify, recognise and extract specific data from it. Data encoded in files that are structured in a machine-readable format should be considered to be machine-readable data. A machine-readable format can be open or proprietary. They can be formal standards or not. Documents encoded in a file format that limits automatic processing, because the data cannot, or cannot easily, be extracted from them, should not be considered to be in a machine-readable format. Member States should, where possible and appropriate, encourage the use of a Union or internationally recognised open, machine-readable format.[36]

Consequently, machine-readable in this context would cover a declaration which is readable for a computer system. Accordingly, machine-readability is only given in the case that the declaration is technical-coded and as a result machine executable.[37]

Recital 35 informs that there is currently no standard for machine readability, emphasising that Member States should use a standard recognised in the Union or internationally. However, the standard does not explicitly advocate a specific standard but simply requires the use of a declaration in machine readable form. Therefore, different Union-wide recognised standards need to be investigated.

---

28 Eleonora Rosati, *Copyright in the Digital Single Market, Article-by-Article Commentary to the Provisions of Directive 2019/790* (OUP 2021) 89.

29 Judgment of the Court (Grand Chamber) of 9 March 2021, *VG Bild Kunst*, C-392/19, ECLI:EU:C:2021:181.

30 Eleonora Rosati, *Copyright in the Digital Single Market, Article-by-Article Commentary to the Provisions of Directive 2019/790* (OUP 2021) 90.

31 Doubting: Niklas Maamar, 'Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen'(2023) ZUM 481, 484; Marcus von Welser, 'Generative KI und Urheberrechtsschranken' (2023) GRUR 516, 519.

32 Malte Baumann, 'Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld' (2023) NJW 3673, 3675.

33 Benjamin Raue, 'Die Freistellung von Datenanalysen durch die neuen Text und Data Mining- Schranken (§§ 44b, 60d UrhG)'(2021) ZUM 793, 795.

34 No member state has implemented a definition of machine-readability into their national transformation of the DSM Directive.

35 <https://dictionary.cambridge.org/dictionary/english/machine-readable> accessed on 04 October 2024.

36 Directive (EU) 2019/1024 European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

37 Malek Barudi, *Das neue Urheberrecht* (1 st edn, Nomos 2021) para. 14.

4.2 Different Concepts to Declare the Opt-Out in a Machine Readable Form

In the following, three different approaches to opt-out declaration in machine readable form are investigated and evaluated.

4.2.1 Declaring the Reservation of TDM Rights through Terms and Conditions

In Recital 18 of the DSM Directive it is stated that: "it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service".

This suggests that opting out may be considered effective if the rightholders reserve their rights in the terms and conditions of a website. However, this overlooks the fact that a website user is not necessarily bound by the terms and conditions or user agreements stored on a website. Furthermore, a declaration in written language is likely not in a machine-readable form as there is no standardized wording that an automated system can verify, and it is unclear which language should be used.

As discussed in Recital 35 of Directive 2019/1024, it can be inferred that "machine-readable" within this context refers specifically to information encoded in computer language that can be processed by automated crawlers. Finally, it is also important to mention that a general reservation such as "all rights reserved" should not be sufficient, as Article 4(3) DSM Directive speaks of an explicit reservation.[38]

4.2.2 Declaring the Reservation of TDM Rights through Robots.txt

Another current practice is the use of so-called robots.txt files.[39]

The robots exclusion standard allows website operators, including search engines such as Google, to recognise whether they are allowed to index the content and display it to their users.[40]

The instruction to exclude a crawler from a website could look like this:

User-agent: *
Disallow: /

However, this has certain disadvantages. Currently, the robots.txt files cannot recognise TDM declarations.[41] In the absence of a TDM declaration, this would not be an explicit reservation, but only an implied reservation, which is not sufficient under the wording of the law.

Another problem is that the use of the robots.txt file can lead to the reserved works no longer appearing in search engines.

This outcome is not desirable, especially since commercial authors rely on being easily discoverable through conventional search engines. A reservation of use in accordance with Article 4(3) of the DSM Directive should not lead to unequal treatment compared to other uses, especially not when displayed as a search engine result. There is therefore a tension between the reservation of use and the interest in being found and listed by search engines. Moreover, it does not align with the legislative intent.[42] However, exceptions can be formulated in the robots.txt file.[43] This means that individual crawlers can be excluded. This can prevent from no longer being listed by the search engine crawlers.

4.2.3 Declaring the Reservation of TDM Rights through TDM Reservation Protocol

This is a proposal by the World Wide Web Consortium in response to a missing definition of machine readability in the DSM Directive.[44]

The objective of this protocol is to enable a rights holder to express their preferences regarding text and data mining of web resources under their control. This facilitates recipients of such declarations to modify their scraping practices accordingly or to negotiate a separate agreement with the rights holder that accommodates all involved parties.[45]

The protocol specifies that the reservation of use is defined as a variable that is assigned the value 0 or 1 by the rights holder. This reservation is already implemented in the HTML source code.

However, the TDM Reservation Protocol is currently still a draft and not an official standard.[46]

4.2.4 Conclusion

Even if certain approaches already exist, they are not mandatory for operators as they are not legally binding.

There remains significant legal uncertainty regarding declaring effectively a reservation of use. Consequently, it is common practice for rights holders to employ multiple standard methods for declaring such reservations concurrently and therefore there is presently often a parallel

---

38 Martin Ebers, Christian A. Heinze, Björn Steinrötter, *Künstliche Intelligenz und Robotik* (1 st edn, CH. Beck 2020) para. 31.

39 Niklas Maamar, 'Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen' (2023) ZUM 481, 484.

40 Ian Peacock, 'Showing Robots the Door, What is Robots Exclusion Protocol?' (1998) <https://ariadne.hosting.lboro.ac.uk/issue/15/robots/> accessed on 04 October 2024.

41 Niklas Maamar, 'Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen' (2023) ZUM 481, 484.

42 Recital 18 of the DSM Directive:" Other uses should not be affected by the reservation of rights for the purposes of text and data mining."

43 David Bomhard, *BeckOK UrhR: UrhG § 44b* (41 st edn, CH. Beck 2024) para. 34.

44 <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdm-rep-20240202/> accessed on 04 October 2024.

45 ibid.

46 The W3C states on their website: "This specification was published by the Text and Data Mining Reservation Protocol Community Group. It is not a W3C Standard nor is it on the W3C Standards Track." <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdm-rep-20240202/> accessed on 04 October 2024.

use of declaring the opt-out in natural language, through robots.txt, as well as through the TDM Reservation Protocol.[47] This is due to the continuing high level of uncertainty among rightholders regarding on how to effectively protect themselves.

It also shows how inefficient the declaration of the opt-out currently is due to the lack of an EU standard.

Introducing an EU standard that is comprehensible and readily accessible is therefore crucial. This would eliminate one of the greatest weaknesses of Article 4(3) of the DSM Directive and lead to an increased legal certainty. The TDM Reservation Protocol is a promising concept in this regard.

To conclude, it would be desirable to see further development by 2026 so that it can be incorporated into the DSM Directive as a common EU standard.

The necessity to declare the opt-out in machine-readable form represents a significant obstacle to the effective application of Article 4(3) of the DSM Directive, as the declaration of the opt-out is inherently associated with considerable uncertainties.

4.3 Lawful Accessibility

Article 4(3) of the DSM Directive states that in order to claim the TDM exception, the content must be "made publicly available online".

Recital 14 of the DSM Directive clarifies that the work must be lawfully accessible. It defines it more clearly regarding cultural heritage institutions and research organizations as "content based on an open access policy or through contractual arrangements", but broadens the scope in the fourth and final sentence of the Recital by stating: "Lawful access should also cover access to content that is freely available online."

The DSM Directive thus only refers to lawful access to the work, but not explicitly to the lawfulness of making the work available.[48] This means that as long as the rightholder does not place his works behind a login or paywall barrier, lawful access can be assumed.[49]

This is where the next problem of effectiveness arises, as training with online piracy sites[50] remains theoretically permitted. Nevertheless, rightholders do not have the ability to opt-out of such platforms, leaving the opt-out mechanism ineffective.[51] The question arises as to whether and how this complexity can be addressed.

Firstly, it might doubtful whether the Directive needs to be amended in the first place.

It can be argued that the requirement is only logical.[52] As analysed, the legislator wanted to promote innovation and thus create legal certainty with creating the TDM exception.

Evidently, the interests of developers were prioritised over the interests of rightholders.

This encounters certain confusion.[53]

For instance, it should not be possible to invoke the TDM exception in the case of obviously illegal websites. Our constitutional state cannot afford to favour an exception via the diversions of illegality. This would contradict a central pillar of European law: the rule of law.

Article 2 TEU[54] states: "The Union is founded on the values of the (…) rule of law (…)".

It is therefore questionable whether the wording in Recital 14 s. 2 is not an editorial mistake by the legislator.[55]

A clear answer cannot be given here, especially as the construct of editorial mistake by the legislator is shaky. However, it is clear that it would make sense to add to Recital 14 that obviously unlawful sources should be excluded from automated data collection. Given the current state of technology, the exclusion of explicitly unlawful pages is possible.[56]

Should there now be calls for important leaked protected subject matter, particularly in the context of journalistic activities, reference should be made to an interpretation in conformity with fundamental rights (Article 11 (2) CFR; Article 10 ECHR), which includes such protected subject matter in the scope of the text and data mining exception that is made accessible by third parties and in the content of which there is a legitimate interest in information that cannot be satisfied in any other way.[57]

This would only minimally interfere with the legislator's aim to foster innovate. In any case, it is questionable to what extent piracy sites can be conducive to innovation, considering that they inhibit innovation by weakening the financial basis of the creative industries, hindering investment, impairing legal markets and infringing intellectual property.

The minimal intervention on the part of developers is offset by a significant improvement in the protection of rightholders. A supplementary amendment or adaptation of the wording would therefore be essential to improve the effectiveness of Article 4 (3) of the DSM Directive.

47 David Bomhard, *BeckOK UrhR: UrhG § 44b* (41 st edn, CH. Beck 2024) para. 38.

48 Thomas Dreier, Gernot Schulze, *Urheberrechtsgesetz: UrhG, ´§ 44b UrhG* (7th edn, CH. Beck 2022) para. 8.

49 Marcus von Welser, ´Generative KI und Urheberrechtsschranken´ (2023) GRUR 516, 518; Malte Baumann, ´Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld´ (2023) NJW 3673, 3675.

50 In this context online piracy sites refer to "The illegal reproduction and distribution of copyrighted material on the Web" <https://www.pcmag.com/encyclopedia/term/internet-piracy> accessed on 04 October 2024.

51 Marcus von Welser, ´Generative KI und Urheberrechtsschranken´ (2023) GRUR 516, 519.

52 David Bomhard, *BeckOK UrhR: UrhG § 44b* (41 st edn, CH. Beck 2024) para. 19.

53 Among others: Niklas Maamar, ´Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen´ (2023) ZUM 481, 485; Malte Baumann, ´Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld´ (2023) NJW 3673, 3675.

54 Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union 2012/C 326/01.

55 *Redaktionsversehen*, BVerwG (German Feder Administrative Court), I B 66.64.

56 Malte Baumann, ´Generative KI und Urheberrecht – Urheber und Anwender im Spannungsfeld´ (2023) NJW 3673, 3675.

57 Benjamin Raue, ´Die Freistellung von Datenanalysen durch die neuen Text und Data Mining- Schranken (§§ 44b, 60d UrhG)´ (2021) ZUM 793, 796.

4.4 Conclusion

At first glance, Article 4(3) of the DSM Directive appears to create a balance between the beneficiaries of the exception and the authors through the opt-out mechanism.

However, this impression proves to be deceptive. Although the intention behind the exception is laudable and offers theoretical the possibility for rightholders to protect both their economic and moral interests in their works, the reality reveals a different picture.

The opt-out mechanism proves to be insufficient for several reasons: there is a lack of a harmonised EU standard for machine-readability to adequately declare the opt-out and the weakness behind the criterion of "lawful accessibility" undermines the whole mechanism.

Furthermore, there is a lack of effective control mechanisms to monitor compliance with the opt-out mechanism and the problems of enforcing rights on the internet, in particular due to the principle of lex loci protectionis in Article 8(1) Rome II, remain.[58]

The question of the adequacy of protection for authors under the current form of Article 4(3) of the DSM Directive can be answered in the negative.

5. THE IMPORTANCE OF CHANGE

5.1 A Pessimistic Outlook for the Future

In recent times, authors have begun to articulate their concerns regarding their perception of AI systems and its developers.[59] They fear AI for various reasons, mainly due to its impact on their creative integrity, livelihood, and artistic rights. There is the fear of losing creative control.

Authors are concerned that AI systems could imitate or even reproduce their individual styles and techniques, leading to a loss of creative uniqueness and control over their work.

Further there is a concern regarding the economic impact.

The use of AI to create artworks may have an impact on the market for original artworks. Reproducible AI-generated artworks is potentially cheaper and more readily available.

Authors fear that this could potentially devalue the work behind it, and therefore their work, and cause them to lose competition with AI in the marketplace.

Authors are also concerned that their work will be used without proper credit and compensation, especially by AI developers and companies using these technologies.

In addition to the economic reasons, there is also an ideological controversy about the use of AI in art: the loss of authenticity and originality. While AI can create aesthetically pleasing works of art, some authors argue that they lack the emotional depth, personal experience and artistic expression of human creativity. The fear is that AI artwork could be seen as equal or even superior, leading to a loss of authenticity and originality in art. This would be a major step backwards for our cultural life.

But there are also questions of ethics, especially the control of technology.

Authors are concerned with the ethical issues surrounding the use of AI in art. The idea that algorithms and data about their work could be used to train or improve AI models raises issues of control, privacy and potential manipulation.

There are concerns among authors that AI could threaten their creative freedom, viability and artistic rights. It is crucial to address these concerns and take the necessary measures to protect the integrity and rights of authors in an increasingly digital world.

As individuals feel powerless against AI, researchers have resorted to innovative ways to outsmart it.

Professor Ben Zhao and his research team at the University of Chicago have developed two tools, Glaze and Nightshade, to protect authors from unwittingly contributing their work to AI training data.[60]

Nightshade manipulates pixels in a way that is imperceptible to humans, but can influence the AI training process by injecting poisoned data into the system, much like a Trojan horse. Manipulated examples can gradually (negatively) influence the entire model.[61]

Data poisoning can lead to the training data being changed in such a way that the model identifies the image of a cow as a horse, for example. Style manipulation is also conceivable, for example the interpretation of an impressionist work as cubist. This can lead to the model delivering inadequate results.[62]

Due to the large amount of training data entered, it is practically impossible for the developers to identify and delete the poisoned data.[63]

It is unfortunate that such drastic measures have to be taken. However, the researchers hope that Nightshade will not only act as a deterrent to AI companies, but also help to strengthen authors' rights and promote a more respectful treatment of their work by putting authors in a stronger position to negotiate with developers.

58 Regulation (EC) No 864/2007 of the European Parliament and of the Council of 11 July 2007 on the law applicable to non-contractual obligations (Rome II).

59 Martin Perhiniak (graphic designer) interviews several authors including Jon Lam, Patrick Brown, Steven Zapata in his documentary "AI vs Authors – The Biggest Art Heist in History "published on his YouTube Chanel <https://www.youtube.com/watch?v=ZJ59g4PV1AE> accessed 04 October 2024.

60 Shawn Shan, Wenxin Ding, Josephine Passananti, Stanley Wu, Haitao Zheng, Ben Y. Zhao, 'Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models' <https://arxiv.org/abs/2310.13828> accessed on 04 October 2024.

61 Melissa Heikkilä, This new Data Poisoning Tool Lets Authors Fight Back Against Generative AI' (2023), MIT Technology Review <https://www.technologyreview.com/2023/10/23/1082189/data-poisoning-authors-fight-generative-ai/> accessed on 04 October 2024.

62 James Thorpe, 'What is Data Poisoning & Why Should You Be Concerned?' (2021), International Security Journal <https://internationalsecurityjournal.com/data-poisoning/> accessed on 04 October 2024.

63 Patrick K. Lin, 'Can This Data Poisoning Tool Help Authors Protect Their Work from AI Scraping?' (2023) Center for art law <https://itsartlaw.org/2023/11/21/can-this-data-poisoning-tool-help-authors-protect-their-work-from-ai-scraping/> accessed on 04 October 2024.

The resistance within the arts industry shows how deeply rooted concerns are about the growing role of AI in creative fields. It highlights a central theme in the current debate about technology and the arts: the importance of preserving humanity and authenticity in an increasingly digital world.

If the opt-out mechanism is not adapted, a bleak picture for the future could be that mechanisms such as Glaze or Nightshade will become authors' preferred means of avoiding scraping, with the consequence that developers will no longer be able to adequately train their AI systems. Consequently, this would also have a negative impact on developers and their AI models, as they are highly dependent on the authenticity and quality of the training data. The risk of "garbage in, garbage out" is undeniable.[64]

Current legal norms are seen by some as too hostile to developers and too copyright-friendly.[65] However, it is clear that the ability of AI to imitate or reproduce artistic works poses new challenges to the integrity and rights of authors.

Resistance to AI in the art community is undeniable, as evidenced by ongoing court cases and the development of tools such as Glaze and Nightshade. These tools are designed to protect authors from unwanted participation in AI training processes and to preserve the integrity of their work. The need for such measures highlights the importance of striking the right balance between technological progress and the protection of artistic rights.

It is essential that legislation such as Article 4(3) of the DSM Directive is adapted accordingly to meet the needs of both authors and developers. This requires close cooperation between legislators, technology companies and their developers, and authors. Mechanisms need to be developed that respect the rights and creative expression of authors while fostering innovation and progress in AI technology.

Overall, the discussion on the role of AI in the arts highlights the need to preserve humanity and authenticity in an increasingly digital world. The development and use of AI should aim to support and enhance creative work without compromising the integrity and rights of authors.

5.2 Forecast: AI Act – Needed Change or Insufficient Block of Innovation?

It is the responsibility of legislators to create a legal framework that ensures a fair balance between developers and authors.

This may not be a utopia in an uncertain future, but an imminent reality due to the AI Act.

Article 53(1)(c) of the AI Act contains the obligation for providers to comply with European copyright law by stating: "put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art

technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790".

While pleasantly, Article 4(3) of the DSM Directive and the opt-out mechanism are explicitly mentioned here, unfortunately, the AI Act fails to define machine readability.

This means that one of the major weaknesses of Article 4(3) of the DSM Directive – the effective declaration of a reservation of use – remains.

Further, Article 53(1) requires providers to "(a) draw up and keep up-to-date the technical documentation of the model, including its training and testing process and the results of its evaluation, which shall contain, at a minimum, the information set out in Annex XI for the purpose of providing it, upon request, to the AI Office and the national competent authorities; (b) draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems. Without prejudice to the need to observe and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law, the information and documentation shall:(i) enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation; and (ii) contain, at a minimum, the elements set out in Annex XII".

It would have been desirable if companies had been obliged to produce detailed summaries in order to create more transparency with regard to training data. However, the high value of trade secrets must be taken into account here. Ultimately, however, confidentiality and transparency are logically mutually exclusive. The legislator has tried to strike a fair balance between the parties here, but the result is an unclear middle ground that does not significantly improve the situation for authors.

However, there is a possible clarification with regard to the tension with the "country-of-origin"-principle in Article 8 II Rome-Regulation and Article 4(3) DSM Directive: Recital 106 of the AI Act states that EU copyright law must be respected in other non-EU countries by saying: "Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place."

In theory, this makes the EU an attractive location for potential developers, as operating within the EU provides a clear legal framework and therefore legal certainty.

In practice, the prognosis seems rather pessimistic.

An import ban on technologies in the digital age is much more difficult to enforce in practice, as the "goods" are not physically imported by ship or plane, but rather unnoticed via the internet.[66] Comprehensive monitoring of internet traffic would fall within the remit of EU cus-

---

64   Katharina de la Durantaye, '»Garbage in, garbage out« – Die Regulierung generativer KI durch Urheberrecht' (2023) ZUM 645, 660.

65   ibid.

66   David Bomhard, Jonas Siglmüller, 'AI Act – das Trilogergebnis' (2024) RDi 45, 46.

toms authorities.[67] There might be also potential constitutional problems with such comprehensive monitoring. For instance, this could violate the protection of personal data (Article 8 CFREU) and the freedom of expression and information (Article 11 CFREU).

Taking this into account, the "import ban" is theoretical good, but the future will show if it is practical and applicable to the "real world".

5.3 Conclusion

It should be noted that the AI Act is one of the first of its kind in the world.[68] The EU has created a legal framework for an important current and future topic relatively quickly. In view of the usually lengthy legislative procedures, this is a positive development.[69]

However, the negative aspects outweigh the positives.

The AI Act is not optimal due to its high level of detail and complexity. This can be seen in the following: while the first draft was 100 pages long, the final draft grew to over 400 pages. Such a detailed legal framework contains the risk that future innovations will be over-regulated. This in its turn is contrary to promoting innovation and the goal of "boosting innovation" as announced in Recital 2. While in the past many processes in the EU were often slow and inefficient due to bureaucracy, it would have been desirable for the AI Act to be less bureaucratic. Unfortunately, it is already failing due to the narrow regulatory framework.

It cannot be assumed that the necessary deep technical understanding is available among the officials involved.[70] Unfortunately, due to its theoretical complexity, the AI Act represents a missed opportunity to promote AI innovation in an appropriate way. It will be interesting to see how it is accepted in practice.

As a result, although in theory an obligation to comply with EU-copyright law and especially with respecting the opt-out is created and an attempt is made to solve the problem of the "country of origin"- principle, in practice this is not very promising.

Without an EU standard for machine readability Article 4(3) of the DSM Directive does not provide adequate protection for authors.

Copyright is indeed a special right for humankind. Its objects are the foundation of our culture. Culture encompasses the entirety of the intellectual, artistic and creative achievements of a community as an expression of human development.[71]

However, the economic aspects should not be underestimated. In modern society, culture has an important economic influence, which can be seen in the entire creative industry landscape. In addition to aesthetic aspects, the visual arts are also essential for a sense of identity and belonging. Both on the part of the authors and on the part of the consumers.

Culture is therefore not only a factor that enriches everyday life, but also an essential component of the development of our society as a whole.

As this article illustrates, there is a growing concern among a number of stakeholders that the rise of AI, and in particular generative AI, poses a threat to the appreciation and further development of cultural assets and the continued existence of cultural life as a whole. In this context, many authors are complaining that their creative output is being devalued by AI. It is therefore essential that intellectual property is adequately protected to the same extent as tangible property.

However, it should be noted that the creative industry has already been repeatedly exposed to technological innovation in the past. The introduction of the camera at the beginning of the 20th century initially posed a threat to the art industry.[72] Over time, however, photography established itself as a significant branch within the art industry.[73] This makes it clear that the application of AI does not necessarily offer great potential for the technology industry, but also for creators who can make use of this new technical tool.

Another current example can be found in the music and film industry. The introduction of online streaming services has presented the music and film industry with new challenges. However, it is evident that these industries have capitalised on the developments with the introduction of streaming services such as Netflix, Apple Music and Spotify.

This demonstrates that the perceived novel dangers associated with generative AI are, in fact, not a recent phenomenon. They are merely happening at an accelerated pace.

As a consequence, it can be stated that AI does not jeopardise our cultural assets and their continued existence or further development.

However, the interests of authors must also be taken into account in such innovations.

Authors are not remunerated for the creative process of their works, but only for the actual utilisation of their work. This ensures that their own income and livelihood

---

67 ibid.

68 The US is currently working on a AI bill of rights, to see latest development: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> accessed on 04 October 2024; China has regulations on AI since 2021 <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117> accessed on 04 October 2024.

69 As mentioned above: This is particularly evident in the 18-year gap between the InfoSoc Directive (2001) and the DSM Directive (2019).

70 David Bomhard, Jonas Siglmüller, 'AI Act – das Trilogergebnis' (2024) RDi 45, 54.

71 John J. Macionis, Linda M. Gerber, *Sociology* (7th edn, PPH 2011) 53.

72 Anthony W. Lee, 'AI or No, It's Always Too Soon to Sound the Death Knell of Art' (2022) <https://www.wired.com/story/art-history-photography-painting-dalle-ai/> accessed on 04 October 2024.

73 This can be seen from the fact that total sales in the photo & video market amounted to around € 11.16 billion in 2022. According to the market forecast, a market volume of € 16.81 billion will be reached in 2027; this corresponds to expected annual sales growth of 9.37% (CAGR 2022–2027). This is only the service sector <https://de.statista.com/outlook/amo/app/foto-video/weltweit>; On the art market, photography has at least a comparable value, for example the photograph "Rhein II" by Andreas Gursky was auctioned for 4.3 million euros <https://design-lovr.de/magazin/fotografie/fotografie-rhein-ii-gursky/> both sources were accessed on 04 October 2024.

are secured. The use of works for AI training purposes constitutes such use. It is therefore only logical that authors should be remunerated accordingly.

There is currently no suitable, EU-wide technical standard for an adequate declaration of the opt-out mechanism. However, this is not an impossible hurdle. A first step would be the introduction of a standardised reservation of use to make the opt-out mechanism practicable feasible.

Although the AI Act provides some clarity, it generally represents a missed opportunity to both promote innovation and adequately protect authors. De facto, authors are currently defenceless against AI developers.

This entails the risk for authors will hide their works behind paywalls in the future, which could lead to a decline in the quality of AI training data in the next step. This situation would undesirable for both sides. It is crucial to establish trust and transparency between the two sides. In the course of the research, it became apparent that a significant number of authors have a negative attitude towards AI, as they fear the loss of human creativity and their work. However, this point of view must be countered by the fact that artificial "creativity" will never come close to human creativity. Instead, it is dependent on it. Without human authors, technical "authors" would not be able to develop further.

In conclusion, it can be stated that Article 4(3) of the DSM Directive does not provide sufficient protection for authors.

Furthermore, in its current version, no balance between the interests of authors and developers can be recognised.

It is to be hoped that the DSM Directive will be re-evaluated adequately in 2026[74] so that the legislator can build a bridge between authors and developers.

The aim should be to use the legal framework to create a world that allows AI and human creativity to co-exist in harmony.
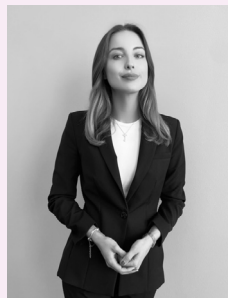
Anna Buss

Anna recently completed her LL.M. in European Intellectual Property Law at Stockholm University, where she dedicated her thesis to examining Article 4(3) of the DSM Directive and its effectiveness (or lack thereof) in protecting artists against the use of their works as AI training data. Previously, she studied law at the University of Mannheim (Germany), graduating with the First Examination in Law. Anna has a particular interest in art law and its further development, especially in the light of the rise of artificial intelligence. She will be starting her legal clerkship in Wiesbaden in November.

---

74   cf. Article 30 of the DSM Directive.