AN EXCEPTIONAL FORMALITY UNDER BERNE: EVASION OF COPYRIGHT PROTECTION VIA THE EU'S TEXT AND DATA MINING EXCEPTION

Larissa Bersh*

ABSTRACT

For more than one hundred years, the international Berne Convention has required signatories to grant copyright protection (1) automatically upon the creation of a work and (2) without subjecting the rights to any "formality." In doing so, the Berne Convention establishes that copyright is inherent to the production of a qualifying work.

The European Union's Digital Single Market Directive 2019/790 turns this long-held principle on its head. Article 4 of the directive, entitled "Exception or limitation for text and data mining," excepts reproductions and extractions for the purposes of text and data mining from copyright protection. Text and data mining describes a family of techniques for machine processing of large volumes of data, and it is central to the value provided by artificial intelligence. Under Directive 2019/790, if an otherwise-copyrighted work is to be used in text and data mining, the protections that authors have long-relied upon may be moot.

This Note analyzes the risks of and potential solutions to this shift. By creating a copyright exception for the use of data for text and data mining, and by requiring authors to take positive action to opt out of the exception, the European Union creates legal uncertainty for both authors and text and data mining developers, violates the Berne Convention, and sets a dangerous precedent for the sanctity and scope of copyright protection in the age of artificial intelligence.

TABLE OF CONTENTS

I. INTRODUCTION

Copyright law is shaped by national interests. In the late nineteenth century, the works of American authors were becoming increasingly popular overseas.[1] In response, after decades of refusal, the United States adopted its first reciprocity-based copyright arrangement to protect American ownership in foreign countries.[2] In the early 2000s, the ability to safely use works with unknown copyright status became necessary to uphold prior legislative actions in the European Union ("EU").[3] In response, the EU enacted a directive excepting certain uses

---

1. *See* Tyler T. Ochoa, *Copyright Protection for Works of Foreign Origin*, *in* THE INTERNATIONALIZATION OF LAW AND LEGAL EDUCATION 167, 171 (Jan Klabbers & Mortimer Sellers eds., Springer Dordrecht 2009) ("It was not until the United States could boast of some authors of international prominence that it finally became in the national interest to extend copyright protection to citizens of other nations on a reciprocal basis."); John A. Rothchild, *How the United States Stopped Being a Pirate Nation and Learned to Love International Copyright*, 39 PACE L. REV. 361, 364 n.4 (2018) ("Until [the middle of the twentieth century], U.S. authors would not have significantly benefited from entering bilateral reciprocity-based copyright treaties or joining the Berne Convention. However, once exports became an important source of income for U.S. authors, it became sensible, and indeed imperative . . . .").

2. *See* Ochoa, *supra* note 1, at 171 ("Finally, in 1891, the U.S. adopted the Chace Act, which extended copyright protection to citizens and residents of foreign nations when those nations agreed to provide copyright protection to U.S. citizens and residents . . . ." (citation omitted)). Note that the United States had long refused to join the Berne Convention, and though this refusal continued for decades more, it joined the Universal Copyright Convention in 1954 "as a stopgap measure." *See* Rothchild, *supra* note 1, at 364 n.4.

3. *See* Agnieszka Vetulani-Cęgiel, *EU Copyright Law, Lobbying and Transparency of Policy-Making*, 6 J. INTELL. PROP., INFO. TECH., & ELEC. COM. L. 146, 155–56 (2015) ("[T]he Commission was in favour of introducing a legal mechanism concerning orphan works as, if no legislation on facilitating rights clearance was adopted, the whole project of the European Digital Libraries Initiative (and Europeana) could fail." (citation omitted)); Council Directive

of "orphan works" from copyright protection.[4] In 2019, as artificial intelligence ("AI") was becoming more efficient,[5] well-funded,[6] and technically advanced,[7] those engaged in text and data mining ("TDM developers") were demanding access to data.[8] In line with the tradition of copyright law, the EU saw an opportunity.[9]

From 1995 to 2010, the quantity of data used to train the leading machine learning models increased by a factor of ten.[10] From 2010 to 2015, it increased by a factor of ten again.[11] From 2015 to 2020, it increased by a factor of 1,000.[12] OpenAI's 2020 GPT-3 model, for instance, was trained on at least 570 gigabytes of data, amounting to nearly one trillion words.[13] Four years later, the leading chatbots were

---

2012/28, recital 1, 2013 O.J. (L 299) 1, 1 (EU) ("[The beneficiaries of this directive] contribute to the preservation and dissemination of European cultural heritage, which is also important for the creation of European Digital Libraries, such as Europeana.").

4. *See* Vetulani-Cęgiel, *supra* note 3, at 152–53 (stating that a directive "on certain permitted uses of orphan works" was issued in 2012); *id.* at 155–56 ("[T]he legislative process on orphan works [was] very much stimulated by the earlier Commission policy actions, i.e. the Digital Libraries Initiative and the Europeana project.").

5. *See, e.g.*, RAYMOND PERRAULT, YOAV SHOHAM, ERIK BRYNJOLFSSON, JACK CLARK, JOHN ETCHEMENDY, BARBARA GROSZ ET AL., ARTIFICIAL INTELLIGENCE INDEX 2019 ANNUAL REPORT 49–50 (Saurabh Mishra ed., Stan. U. 2019), https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf [https://perma.cc/5MAA-RU9B] (showing that the training time and cost of an image classification model decreased since 2017).

6. *See id.* at 88 ("Globally, investment in AI startups continues its steady ascent.").

7. *See id.* at 51–56 (showing that AI capabilities in image generation, image recognition and natural language understanding were improving significantly by 2019).

8. *See* Filiz Ersoz, *Data Mining and Text Mining with Big Data: Review of Differences*, 6 INT'L J. RECENT ADVANCES MULTIDISCIPLINARY RSCH. 4391, 4392 (2019) ("To be able to perform data mining; access to data . . . [is] required."); Rossana Ducato & Alain Strowel, *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility*,*"* 50 INT'L R. INTELL. PROP. & COMPETITION L. 649, 651 (2019) ("Many studies have already highlighted the need for a broad access to datasets so as to train algorithms and improve AI applications.").

9. *See infra* notes 21–23 and accompanying text (indicating the EU's interest in developing an AI strategy).

10. *See* Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn & Pablo Villalobos, *Compute Trends Across Three Eras of Machine Learning*, PROC. 2022 INT'L JOINT CONF. NEURAL NETWORKS, at 3 (2022) (showing that training dataset sizes increased from roughly 10^5 datapoints in 1995 to 10^6 datapoints in 2010 during the "Pre Deep Learning Era").

11. *See id.* (showing that training dataset sizes increased from roughly 10^6 datapoints in 2010 to 10^7 datapoints in 2015 during the "Deep Learning Era")*.*

12. *See id.* (showing that the training dataset sizes for "large scale" models increased from roughly 10^8 datapoints in 2015 to 10^{11} datapoints in 2020 during the "Large Scale Era")*.* This exponential growth has only continued. *Id.* (showing multiple datasets in 2021 exceed 10^{12} datapoints)*.*

13. *See* Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal et al., *Language Models are Few-Shot Learners*, PROC. 34TH CONF. ON NEURAL INFO. PROCESSING SYS., at 8 (2020).

trained on as many as three trillion words.[14] In short, the creation of AI models depends on human-authored works[15] — and a lot of them. The problem is that some of these works are protected by copyright.[16]

Before 2019, some text and data mining ("TDM") uses were swept under the umbrella of existing copyright exceptions under EU law. For example, data reproduced for use in TDM may comprise transient copies, and thus may be excepted from the exclusive right of reproduction under the Copyright and Information Society Directive ("InfoSoc Directive").[17] But many TDM uses were not explicitly covered by such

---

14. *See* Cade Metz, Cecilia Kang, Sheera Frenkel, Stuart A. Thompson & Nico Grant, *How Tech Giants Cut Corners to Harvest Data for A.I.*, N.Y. TIMES (Apr. 8, 2024), https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelli gence.html [https://perma.cc/WNA8-WPJE].

15. *See* Daniel Rodríguez Maffioli, Copyright in Generative AI Training: Balancing Fair Use through Standardization and Transparency 4 (Aug. 21, 2023) (unpublished manuscript) (on file with SSRN), https://ssrn.com/abstract=4579322 [https://perma.cc/S84F-PKKV] (referring to the "copious amounts of copyrighted data gathered for training [AI] models"). It is generally accepted that copyrighted work under the Berne Convention requires human authorship. AI models' dependence on copyrighted work is thus a dependence on human-authored work. *See* Sam Ricketson, *The 1992 Horace S. Manges Lecture — People or Machines: The Berne Convention and the Changing Concept of Authorship*, 16 COLUM.-VLA J.L. & ARTS 1, 8, 11 (1991) (laying out textual and other evidence to support the argument that "author" in the Berne Convention refers to "a human creator"). Developers' need for data has motivated research into "synthetic data," or data created by computers. *See* Ruibo Liu, Jerry Wei, Fangyu Liu, Chenglei Si, Yanzhe Zhang, Jinmeng Rao et al., *Best Practices and Lessons Learned on Synthetic Data for Language Models*, 2024 CONF. ON LANGUAGE MODELLING, at 1 (2024) (hailing synthetic data as a "promising solution" to address data scarcity and cost). However, synthetic data carries significant limitations. *See id.* at 2 (listing scholarship on synthetic data's shortcomings, which include lost factuality and fidelity, amplified biases, and new biases). Indeed, the insufficiency of artificial data is readily apparent in how aggressively developers seek out human-created data. *See* Metz et al., *supra* note 14 (describing how tech companies have attempted to access copyrighted data for use in training their AI models, including the transcription of YouTube videos and discussions regarding the possible purchase of a publishing house).

16. *See* Kalpana Tyagi, *Copyright, Text & Data Mining and the Innovation Dimension of Generative AI*, 19 J. INTELL. PROP. L. & PRAC. 557, 563 ("The entire value chain of GenAI is copyright-driven. Consider for instance the input, that is the content, and datasets that are used to train these models." (citations omitted)); Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 745 ("Creating a training set of millions of examples almost always requires, first, copying many more millions of images, videos, audio, or text-based works. Those works are almost all copyrighted" (citation omitted)). Note that Lemley & Casey is based on United States law; however, as both the U.S. and EU are subject to the Berne Convention, the examples listed are works that are generally subject to copyright protection under EU law as well. *See* Päivi Hutukka, *Copyright Law in the European Union, the United States and China*, 54 *Int'l Rev. Intell. Prop. & Competition L.* 1044, 1069 (stating similarities in EU and US copyright law are due in part to shared obligations under the Berne Convention); Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, S. TREATY DOC. No. 99-27 [hereinafter Berne Convention], at art. 2(1) (providing copyright protection to "literary and artistic works" including books, cinematographic works, drawings, photographic works, and more).

17. *See* Council Directive 2001/29, 2001 O.J. (L 167) 1, 16 [hereinafter InfoSoc Directive], at art. 5(1) (excepting from copyright "[t]emporary acts of reproduction . . . which are transient or incidental [and] an integral and essential part of a technological process and whose sole

exceptions.[18] This fact, in addition to the optional nature of most of the existing exceptions leading to unharmonized results across states,[19] meant that significant legal uncertainty remained.[20]

Fortunately for TDM developers, concurrent with their increasing need for access to copyrighted data, the EU announced its ambitions "to become a leader in the AI revolution."[21] While some countries had moved forward on AI strategy and investment, the EU had fallen behind,[22] and it was quickly realizing that attracting AI industry players to the EU could be important.[23] In 2019, the EU acted. This took the form of the 2019 Digital Single Market Directive (the "DSM

---

purpose is to enable . . . a lawful use of a work . . . and which have no independent economic significance" (alteration in original)); The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market — Legal Aspects, Rep. to the European Parliament Policy Department for Citizens' Rights & Constitutional Affairs, PE 604.941, at 9 (2018) [hereinafter Legal Affairs Analysis] ("The [InfoSoc Directive's] mandatory exception for temporary acts of reproduction might apply to limited TDM techniques."); Council Directive 2019/790, 2019 O.J. (L 130) 92, 93 (EU) [hereinafter DSM Directive], at recital 18 (citing the InfoSoc Directive as an exception under which TDM uses may fall).

18. Legal Affairs Analysis, *supra* note 17, at 9 (stating that pre-DSM Directive, "TDM might be possibly covered by exceptions and limitations available, however their application is uncertain" (emphasis removed)); *id.* at 9–12 (listing various potential exceptions under which TDM activities might fall, subject to limitations and/or legal uncertainty); *id.* at 12 ("[I]t is highly uncertain whether existing exceptions and limitations, both mandatory and voluntary, would apply to TDM.").

19. *See* Legal Affairs Analysis, *supra* note 17, at 12 ("[I]t should be noted that all mentioned exceptions and limitations that could apply to TDM — but the exception of temporary acts of reproduction — are implemented by Member States on a voluntary basis. Voluntary implementation makes even less predictable whether existing exceptions and limitations can be applied to TDM projects, especially those of cross border nature.").

20. *See id.* (connecting researchers' legal uncertainty regarding the legality of their TDM uses with the uncertain application of existing exceptions and the legal fragmentation caused by optional implementation across Member States); DSM Directive, *supra* note 17, at recital 3 (stating that as of the writing of the DSM Directive, copyright owners and users faced remaining legal uncertainty regarding "certain uses" of copyrighted works, including cross-border uses); *id.* at recital 18 ("[U]sers of text and data mining could be faced with legal uncertainty . . . in particular when the reproductions or extractions made for the purposes of the technical process do not fulfil all the conditions of the existing exception for temporary acts of reproduction provided for in [the InfoSoc Directive].").

21. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, at 19, COM (2018) 237 final (Apr. 25, 2018) (emphasis omitted).

22. *See id.* at 4 (citing the AI strategies and investment figures of the United States and China before stating that Europe is "behind" compared to Asian and North American private AI investment and that "[o]nly a fraction of European companies have already adopted digital technologies").

23. *See id.* at 11 ("Finally, the EU needs to train more specialists in AI, building on its long tradition of academic excellence, create the right environment for them to work in the EU and attract more talent from abroad." (emphasis omitted)).

Directive"),[24] legislation creating exceptions to the ordinary copyright scheme outlined in the international Berne Convention.[25]

The DSM Directive creates two mandatory exceptions to copyright related to TDM: Article 3 and Article 4.[26] Article 3 is entitled "Text and data mining for the purposes of scientific research" and requires member states to allow "reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining."[27]

While the Article 3 exception creates a clear carve-out from the exclusive right of reproduction, the Article 4 exception is more ambiguous. Titled "Exception or limitation for text and data mining," Article 4 requires member states to adopt an exception for "reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining."[28] This exception broadly applies to all TDM activities, including the development of large language models ("LLMs"), which depend on TDM of "large amounts of data" in the training process.[29] Article 4's application is limited only by a controversial feature: it allows rightholders to reserve use of their

24. *See id.* at 2 (listing "the Digital Single Market" as part of the EU's approach "to make the most of the opportunities offered by AI"); *see* DSM Directive, *supra* note 17, at recital 1 (embodying the strategy of a digital single market through "[h]armonisation of the laws of the Member States on copyright and related rights"); Hanjo Hamann, *Artificial Intelligence and Machine-Readability: A Review of Human-to-Machine Communication Protocols and their (In)Compatibility with Article 4(3) of the Copyright DSM Directive*, 15 J. INTELL. PROP., INFO. TECH., & ELEC. COM. L. 102, 106 (2024) ("[T]he TDM exception is meant 'to encourage innovation also in the private sector' through incentivizing AI developers.").

25. *See* DSM Directive, *supra* note 17, at recital 11 (creating an exception to the right of reproduction for certain entities, including research organizations, engaging in text and data mining); *id.* at recital 18 (creating an exception "under certain conditions" to the right of reproduction when works are used for text and data mining); Berne Convention, *supra* note 16, at art. 9(1) ("Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works."). Note that directives under EU law do not automatically apply to Member States; they must be transposed by each country. *See* Publications Office of the European Union, *Directive*, EUR-LEX (n.d.), https://eur-lex.europa.eu/EN/legal-content/glossary/directive.html [https://perma.cc/FGT4-58UN]. Here, Article 4 represents a "mandatory minimum," meaning that Member States are required to implement the exception, may not minimize the scope of the exception, and indeed may "allow[] greater scope for exempted use." European Copyright Society, 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 (May 3, 2022), at 2, https://europeancopyrightsociety.org/wp-content/uploads/2022/05/ecs_exceptions_final-3.pdf [https://perma.cc/L9SJ-TLSS].

26. *See* Thomas Margoni & Martin Kretschmer, *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology*, 71 GRUR INT'L 685, 685 (2022); DSM Directive, *supra* note 17, at art. 3–4.

27. DSM Directive, *supra* note 17, at art. 3 & ¶ 1.

28. *Id.* at art. 4 & ¶ 1.

29. Tyagi, *supra* note 16, at 562 (2024). *See also* Helen Toner, *What Are Generative AI, Large Language Models, and Foundation Models?*, CTR FOR SEC. & EMERGING TECH. (May 12, 2023), https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/ [https://perma.cc/KH47-45V4] (explaining that large language models are a subset of generative AI systems).

works.[30] Article 4(3) states that the exception applies so long as the use of the works in question "has not been expressly reserved by their rightholders."[31] This requires creators to opt *into* copyright protection via a reservation of their rights rather than opt *out* of it.[32]

For over one hundred years, it has been the policy of the Berne Convention that authors receive automatic copyright protection under international law.[33] Berne accomplishes this by prohibiting its members, including the EU,[34] from imposing formalities on the protection of covered works. That is, copyright protection cannot be dependent on conditions such as whether the author registered[35] or provided public notice of their work.[36] Instead, member states are directed to provide foreign authors with copyright protection immediately and without qualification[37] upon a work's embodiment, "whatever may be the mode

---

30. *See* DSM Directive, *supra* note 17, at art. 4(3) (allowing the exception to apply to a work only if it "has not been expressly reserved" by the work's rightholder); Hamann, *supra* note 24, at 107 (suggesting that there are "[t]wo camps" with "diametrically opposing fears" regarding Article 4's reservation right).

31. *Id.* at art. 4, ¶ 3.

32. *See* Séverine Dusollier, *(Re)introducing Formalities in Copyright as a Strategy for the Public Domain*, *in* OPEN CONTENT LICENSING: FROM THEORY TO PRACTICE 75, 76 (Lucie Guibault & Christina Angelopoulos eds., 2012) (characterizing formalities that act "as gateways through which the creator should pass in order to benefit from the protection of copyright" as "opt-in mechanisms" (emphasis omitted)).

33. *See Berne Convention for the Protection of Literary and Artistic Works (Berne Convention)*, THOMSON REUTERS PRAC. L., https://us.practicallaw.thomsonreuters.com/3-502-8945 [https://perma.cc/SK9B-USWH] [hereinafter *Berne Convention Guide*].

34. As members of the World Trade Organization ("WTO"), the EU is a party to the Agreement on Trade-Related Aspects of Intellectual Property Rights ("TRIPS"). *See The EU and the WTO*, EUR. COMM'N, https://policy.trade.ec.europa.eu/eu-trade-relationships-country-and-region/eu-and-wto_en [https://perma.cc/B84B-CKZY] ("Both the European Union (EU) and the individual EU countries are members of the WTO"); *Frequently asked questions about TRIPS [trade-related aspects of intellectual property rights] in the WTO*, WORLD TRADE ORG., https://www.wto.org/english/tratop_e/trips_e/tripfq_e.htm [https://perma.cc/P5PL-7TJ6] ("The TRIPS Agreement . . . applies to all WTO members."). The agreement incorporates the obligations of the Berne Convention in Article 9.1. *See* Agreement on Trade-Related Aspects of Intellectual Property Rights part II section 1 art. 9, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299 [hereinafter TRIPS Agreement]; *see also Overview: the TRIPS Agreement*, WTO https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm [https://perma.cc/WQS8-BHRC] (explaining that TRIPS "requir[es] compliance with the basic standards" of Berne and adds enforcement and dispute settlement provisions); Paulien Wymeersch, *EU Copyright Exceptions and Limitations and the Three-Step Test: One Step Forward, Two Steps Back*, 72 GRUR INTERNATIONAL 631, 636 (2023) ("Even though the EU as an organization is not a party to the Berne Convention, the CJEU has declared that the Union is nevertheless obliged to comply with . . . the Berne Convention.").

35. *See Berne Convention*, CORNELL U. LEGAL INFO. INST. (Nov. 2021), https://www.law.cornell.edu/wex/berne_convention [https://perma.cc/QHQ5-CA23] ("[G]enerally, members [of the Berne Convention] must not require registration of works from foreign citizens.").

36. *See* Berne Convention Implementation Act of 1988, 17 U.S.C. § 401 (removing the requirement to provide notice).

37. *See Berne Convention Guide*, *supra* note 33.

or form of its expression."[38] These protections include the exclusive rights of translation,[39] reproduction,[40] and performance[41] of one's work, among others.

Automatic protection allows authors to avoid the "cumbersome, costly, and often unsuccessful" process of compliance with the formalities of foreign countries.[42] If Berne's "formalities" did not extend to registration, for example, then authors could be required to manually opt into protection for each country in which they want ownership of their own work.[43] Each country could have its own procedure for registration, with different requirements, timelines, and expenses.[44] The economic[45] and Hegelian[46] motivators of copyright law would be marred by bureaucracy, transaction costs, and self-interested national incentives.[47]

With that being said, the automatic rights granted by Berne are not without exceptions,[48] including those outlined in Articles 3 and 4 of the

---

38. *See* Berne Convention, *supra* note 16, at art. 2, ¶ 1. Individual member states may choose to create a fixation requirement, such that works must be "fixed in some material form" in order to be eligible for protection. Nicola Lucchi, *Genetic Copyright: An Alternative Method for Protecting and Using Essential Public Knowledge Assets?*, 40 EUR. INTELL. PROP. REV. 766, 768 (2018) (quoting Berne Convention, *supra* note 16, at art. 2, ¶ 2).

39. *See* Berne Convention, *supra* note 16, at art. 8.

40. *See id.* at art. 9.

41. *See id.* at art. 11.

42. *See* Jane C. Ginsburg, *Berne-Forbidden Formalities and Mass Digitization*, 96 B.U. L. REV. 745, 749 (2016).

43. *See* Brad A. Greenberg, *More Than Just a Formality: Instant Authorship and Copyright's Opt-Out Future in the Digital Age*, 59 UCLA L. REV. 1028, 1030, 1031 (2012) (explaining that, where formalities like registration apply, "copyright only vest[s] in a work if its author satisfie[s]" those formalities in a regime that has been referred to as an "opt-in copyright system").

44. *See* Dev S. Gangjee, *Copyright Formalities: A Return to Registration?*, *in* WHAT IF WE COULD REIMAGINE COPYRIGHT? 213, 216 (Rebecca Giblin & Kimberlee Weatherall eds., 2017) (stating that Berne's ban on formalities "responded to the difficulties experienced by creators confronted with a daunting array of jurisdiction-specific formalities").

45. *See* Mitchell Longan, *A System Out of Balance: A Critical Analysis of Philosophical Justifications for Copyright Law Through the* Lenz *of Users' Rights*, 56 U. MICH. J.L. REFORM 779, 794–801 (2022).

46. *See* Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287, 330–50 (1988).

47. Individual nations value the economic success of their own citizens and businesses. *See, e.g.*, Rothchild, *supra* note 1, at 364 (discussing the United States' failure to protect foreign works where Americans were benefitting). Absent international agreement, this can incentivize them to create excessive hurdles for foreign authors to acquire protection. *See id.* at 365–66 (discussing the protectionist "manufacturing clause" included in U.S. copyright law).

48. Article 13 of TRIPS contemplates that member states may create "limitations or exceptions to exclusive rights." TRIPS Agreement, *supra* note 34, at art. 13. For example, the InfoSoc Directive legislates that "[t]emporary" or "transient" acts of reproduction must be excepted from the right of reproduction. *See* InfoSoc Directive, *supra* note 17, at art. 5, ¶ 1. Such acts are an integral part of various technological processes, including "caching and browsing." *See* STAVROULA KARAPAPA, *Transient Uses and Temporary Copying*, *in* DEFENCES TO COPYRIGHT INFRINGEMENT: CREATIVITY, INNOVATION AND FREEDOM ON THE

DSM Directive.[49] But Berne limits all exceptions with certain constraints. When EU member states legislate limitations or exceptions to the right of reproduction, Article 9(2) of Berne requires them to restrict those exceptions to "[(1)] certain special cases, provided that [(2)] such reproduction does not conflict with a normal exploitation of the work and [(3)] does not unreasonably prejudice the legitimate interests of the author."[50] This requirement in the Berne Convention is known as the "three-step test" defining the permissibility of an exception or limitation on copyright. While the first step requires an exception to be "clearly defined" and "narrow in its scope and reach,"[51] steps two and three broadly focus on an exception's economic impact.[52]

Despite its importance, the three-step test has seen limited practical analysis.[53] June 2000 was "the first time an international adjudicative body has interpreted" the test, in a World Trade Organization ("WTO")

INTERNET 112, 112 (2020). Unlike the DSM Directive, this exception is narrow and author-protective, requiring reproductions to be "transient or incidental [], forming an integral and essential part of a technological process and carried out for the sole purpose of enabling either efficient transmission in a network . . . or a lawful use of a work" and without "separate economic value on their own." InfoSoc Directive, *supra* note 17, at recital 33. The InfoSoc Directive contains other exceptions, including certain uses of copyrighted work for "teaching or scientific research," though unlike the exception for transient acts of reproduction, the remaining exceptions are optional for member states. *Id.* at art. 5, ¶ 3.

49. *See* Berne Convention, *supra* note 16, at art. 9 (granting authors the exclusive right of reproduction); see DSM Directive, *supra* note 17, at arts. 3–4 (providing exceptions for the reproduction of works used for TDM); *see also* Wymeersch, *supra* note 34, at 636 ("Even though the EU as an organization is not a party to the Berne Convention, the CJEU has declared that the Union is nevertheless obliged to comply with . . . the Berne Convention.").

50. Berne Convention, *supra* note 16, at art. 9, ¶ 2. This language is incorporated into Article 13 of the TRIPS Agreement, making it enforceable in the EU and expanding its coverage from the right of reproduction to all Berne-protected copyrights. *See* TRIPS Agreement, *supra* note 34, at art. 13; Panel Report, United States — Section 110(5) of the US Copyright Act, at 15, WTO Doc. WT/DS160/R (Jun. 15, 2000), https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=Q:/WT/DS/160R-00.pdf&Open=True [https://perma.cc/J6HC-MHPT] [hereinafter DS160 Panel Report]. Though Berne's version of the three steps focuses on individual countries' applications of the test, the TRIPS Agreement focuses on members' obligation to follow the test. *See* Berne Convention, *supra* note 16, at art. 9, ¶ 2; TRIPS Agreement, *supra* note 34, at art. 13. The EU's status as a member of the TRIPS Agreement subjects it to this obligation. *See The European Union and the WTO*, WORLD TRADE ORG., https://www.wto.org/english/thewto_e/countries_e/european_communities_e.htm [https://perma.cc/JMA4-3NTP] (listing the EU as a member of the WTO, and thereby as a member of TRIPS).

51. DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(ii), ¶ 6.112.

52. *See id.* at § VI(D)(2)(c)(i), ¶ 6.183 (considering whether users under the exception "enter into economic competition with the ways that right holders normally extract economic value from that right to the work"); *id.* at § IV(D)(1)(d)(i), ¶ 6.229 (considering the "loss of income to the copyright owner" caused by the exception).

53. *See* María Vasquez Callo-Müller, *FTAs' Contribution Towards a More Flexible Copyright Space: Possibilities and Limits*, 38 AM. U. INT'L L. REV. 159, 184 (2023) ("[T]he only time the WTO dispute settlement resolution interpreted the test was . . . more than 20 years ago. Moreover, this is the only time that an international tribunal interpreted the test." (footnotes omitted)).

dispute resolution panel decision.[54] The panel's analysis confirmed that each condition is "a separate and independent requirement that must be satisfied" for an exception or limitation to be permissible and applied "[t]he principle of effective treaty interpretation" to give meaning to each step.[55]

The DSM Directive's Article 4 exception has brought the three-step test back to the forefront of contemporary copyright law — especially now, as there are enormous economic and political pressures from AI-assisted industries to push the envelope further.[56] As the AI industry continues to grow,[57] the DSM Directive may herald a new vision for creative expression and human ownership in the coming century. It is critical that policymakers appreciate that the DSM Directive is a policy shift from the Berne Convention[58] and carefully consider whether that shift is a good one.

This Note argues that the DSM Directive creates unpredictable legal results, violates two articles of the Berne Convention, and sets the

---

54. Jane C. Ginsburg, *Toward Supranational Copyright Law? The WTO Panel Decision and the "Three-Step Test" for Copyright Exceptions* 3 (Ctr. L. & Econ. Stud., Working Paper No. 181, 2001).

55. DS160 Panel Report, *supra* note 50, at § 2(a), ¶ 6.97.

56. *See* Billy Perrigo, *OpenAI Could Quit Europe Over New AI Rules, CEO Sam Altman Warns*, TIME (May 25, 2023), https://time.com/6282325/sam-altman-openai-eu/ [https://perma.cc/K4K2-63EN] ("OpenAI CEO Sam Altman said Wednesday his company could 'cease operating' in the European Union if it is unable to comply with the provisions of new artificial intelligence legislation that the bloc is currently preparing."); Adam Satariana & Cecilia Kang, *How Nations Are Losing a Global Race to Tackle A.I.'s Harms*, N.Y. TIMES (Dec. 6, 2023), https://www.nytimes.com/2023/12/06/technology/ai-regulation-policies.html [https://perma.cc/6FB3-WS4S] ("[T]ech executives warned that overly aggressive regulations could put Europe at an economic disadvantage."); Billy Perrigo, *Big Tech Is Already Lobbying to Water Down Europe's AI Rules*, TIME (Apr. 21, 2023), https://time.com/6273694/ai-regulation-europe/ [https://perma.cc/AZ2R-V46W] (detailing tech companies' lobbying efforts to exclude general purpose AI models from certain AI Act regulations after spending "billions" on AI).

57. *See* Press Release, Bloomberg, *Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds* (June 1, 2023), https://www.bloomberg.com/company/press/genera tive-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/ [https://perma.cc/P626-5GKQ].

58. In the entirety of the EU's Proposal, Opinions, and First Reading — including over 35 Discussions within the Council or its preparatory bodies — there is only one relevant mention of the Berne Convention. *See Commission Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM (2016) 593 final (Sept. 14, 2016) (making no mention of the Berne Convention); Letter from Roberto Viola to Pavel Svoboda, Chair of the Legal Affairs Committee, European Parliament and Neil Kerr, Ambassador and Deputy Permanent Representative, Committee of the Permanent Representatives of the Governments of the Member States to the European Union (Apr. 27, 2017), at 7, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_8508_2017_INIT [https://perma.cc/2JPU-RJRD] (sole mention of the Berne Convention by stakeholders). In this sole mention, a stakeholder "proposed to introduce a flexible norm based on the three-step test in the Berne Convention as a way to future-proof exceptions." Letter from Roberto Viola to Pavel Svoboda, *supra*, at 7. The EU received input from different stakeholders suggesting that the TDM exception should be narrower in scope, but the Berne Convention was not mentioned in this critique. *Id.* at 5 ("[The stakeholders] considered that the new exception should be limited to non-commercial purposes and advocated for a narrow scope.").

stage for future departures from Berne-abiding practices. Part II argues that Article 4 creates legal uncertainty for both rightholders and TDM developers. Part III contends that Article 4 violates the Berne Convention on two grounds. Part IV explores the dangerous precedent set by Article 4. Part V offers several solutions to the problems posed by Article 4, including through the EU's 2024 Artificial Intelligence Act[59] and amendments to Article 4.

II. ARTICLE 4 OF THE DSM CREATES LEGAL UNCERTAINTY

While the Berne Convention is marked by an aversion to formalities,[60] the DSM Directive seems to bemoan their absence. In particular, the DSM Directive's recitals[61] rue the legal uncertainty they claim is caused in part by the need to obtain approval from rightholders before mining copyrighted content.[62] Articles 3 and 4 of the DSM Directive are intended to address this uncertainty.[63] In reality, Article 4 compounds it.

Article 3 creates an exception from copyright protection when "reproductions and extractions [are] made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining."[64] Under this provision, "research organisations" are non-profit or public-interest oriented entities with the primary goal of conducting scientific research.[65] "Cultural heritage institutions" include publicly accessible libraries, museums,

---

59. Council Regulation 2024/1689 of Jun. 13, 2024, Artificial Intelligence Act, 2024 O.J. (L) 1 [hereinafter AI Act].

60. *See* Berne Convention, *supra* note 16, at art. 5, ¶ 2 ("The enjoyment and the exercise of these rights shall not be subject to any formality . . . .").

61. Recitals refer to "[t]ext at the start of an EU act that sets out the reasons for its operative provisions"; they are not legally binding, but "can be important in interpreting [an] ambiguous provision." Thomson Reuters, *Recital (EU)*, PRACTICAL L., https://uk.practicallaw.thomson reuters.com/w-009-6368 [https://perma.cc/W953-GCFH].

62. *See* DSM Directive, *supra* note 17, at recital 8 ("[Research organisations and cultural heritage institutions] are confronted with legal uncertainty as to the extent to which they can perform text and data mining of content . . . . Where no exception or limitation applies, an authorisation to undertake [acts protected by copyright] is required from rightholders.").

63. *See id.* at recital 11 ("The legal uncertainty concerning text and data mining should be addressed by providing for a mandatory exception for universities and other research organisations, as well as for cultural heritage institutions, to the exclusive right of reproduction and to the right to prevent extraction from a database."); *id.* at recital 18 ("In order to provide for more legal certainty . . . this Directive should provide, under certain conditions, for an exception or limitation for reproductions and extractions of works or other subject matter, for the purposes of text and data mining . . . .").

64. *Id.* at art. 3, ¶ 1.

65. *Id.* at recital 12.

archives, and other similar entities.[66] The exception "applies both to commercial and non-commercial uses."[67]

The DSM Directive justifies Article 3 by claiming that cultural heritage and research organizations face "legal uncertainty as to the extent to which they can perform text and data mining of content."[68] The recitals detail potential sources of legal uncertainty including the scope of pre-DSM Directive exceptions;[69] their applicability in different member states;[70] the copyrightability of the underlying work;[71] and the possibility that organizations may be required to receive rightholder authorization before using content for TDM.[72]

In the face of this uncertainty, a data miner's safest bet to ensure legal compliance may be to treat rightholder authorization as a de facto requirement.[73] Article 3's copyright exception responds to this by removing any possibility that rightholder authorization is required.[74] In doing so, the exception harmonizes the legality of TDM for research and cultural heritage institutions in the EU.[75]

---

66. *Id.* at recital 13.

67. Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU* 28 (manuscript at 28) (Jan. 21, 2020) (Ctr. for Int'l Intell. Prop. Stud., Research Paper No. 2019-08) (on file with SSRN), https://ssrn.com/abstract=3470653 [https://perma.cc/PBE2-LMFD]; *see also* Maria-Daphne Papadopoulou, Krystallenia Kolotourou & Maria Bottis, *The Exception of Text and Data Mining from the Academic Libraries Standpoint*, 9 OPEN J. SOC. SCIS. 502, 507 (2021) ("[I]t is conceived as entailing both commercial and non-commercial uses. However, the said inclusion of both commercial and non-commercial uses should not be confused with the character of the beneficiaries of this exception since these are explicitly limited to non-for-profit research organizations and to publicly accessible cultural heritage institutions.").

68. DSM Directive, *supra* note 17, at recital 8. Note that legislators were not motivated by legal uncertainty alone in crafting this exception. *See id.* at recital 10 (expressing concern "that the Union's competitive position as a research area will suffer, unless steps are taken to address the legal uncertainty concerning text and data mining.").

69. *See id.* at recital 18 ("[U]sers of text and data mining could be faced with legal uncertainty as to whether reproductions and extractions . . . can be carried out . . . when . . . [they] do not fulfil all the conditions of the existing exception for temporary acts of reproduction.").

70. *See id.* at recital 10 ("Union law provides for certain exceptions and limitations covering uses for scientific research purposes which may apply to acts of text and data mining. However, those exceptions and limitations are optional.").

71. *See id.* at recital 9 ("Text and data mining can also be carried out in relation to mere facts or data that are not protected by copyright, and in such instances no authorisation is required under copyright law.").

72. *See id.* at recital 8 ("Where no exception or limitation applies, an authorisation to undertake such acts is required from rightholders.").

73. *See, e.g.*, William M. Hannay, *Legally Speaking: Of Mindfields and Minefields: Legal Issues in Text and Data Mining*, 26 AGAINST THE GRAIN 52, 54 (2014) (suggesting that in the face of legal uncertainty under U.S. law, "the conservative approach" to accessing copyrighted material for the purpose of mining was "to seek permission from the owner of the protected works").

74. *See* DSM Directive, *supra* note 17, at recital 11 ("The legal uncertainty concerning text and data mining should be addressed by providing for a mandatory exception for universities and other research organisations, as well as for cultural heritage institutions, to the exclusive right of reproduction and to the right to prevent extraction from a database.").

75. *See id.*

This course of action is not completely without merit. One can indeed imagine that researchers faced uncertainty, not the least of which stemmed from the potential need for rightholder authorization.[76] After all, the substance of an authorization may be less than straightforward,[77] and TDM developers are charged with correctly understanding, acting on, and in some cases, monitoring changes in these permissions.[78] A misunderstanding of the contours of an authorization may lead to copyright infringement for the TDM developer.

But ironically, while removing rightholder authorization may theoretically create legal certainty under Article 3, the implementation in Article 4 creates the opposite effect. At first, Article 4(1) seemingly succeeds in removing rightholder authorization by creating a general exception to the protection of copyrighted work for TDM uses.[79] But Article 4(3) essentially undoes this by creating an authorization requirement of its own.[80] Though Article 4(3) is worded as authors' ability to "reserve[]" copyright protection of their work such that they opt out of their work's use for TDM,[81] it functions identically to an authorization requirement. After all, if Article 4(3)'s reservation function is to have

---

76. *See* Maria Kanellopoulou-Botti, Marinos Papadopoulos, Christos Zampakolas & Paraskevi Ganatsiou, *Legal and Technical Issues for Text and Data Mining in Greece*, 2019 COMPUT. ETHICS — PHIL. ENQUIRY PROC. (2019) 1, 8 n.6, https://digitalcom mons.odu.edu/cgi/viewcontent.cgi?article=1018&context=cepe_proceedings [https://perma.cc/3GQF-7X2S] ("Research organizations and researchers do not always know whether TDM is copyright-relevant at all, whether it may be covered by an exception or whether a specific rightholders' authorization is required.").

77. *See* Hamann, *supra* note 24, at 114 (surfacing the possibility that a reservation attempts to "communicate conditional permissions, as would be needed to reserve TDM content for automatable commercial licensing," a subtlety that some forms of machine-readable reservation are not equipped to express).

78. *See* Lisa Lobling, Christian Handschigl, Kai Hofmann & Jan Schwedhelm, *Navigating the Legal Landscape: Technical Implementation of Copyright Reservations for Text and Data Mining in the Era of AI Language Models*, 14 J. INTELL. PROP., INFO. TECH., & ELEC. COM. L. 499, 501 (2024) ("The TDM user bears the onus of proof . . . . Thus, the user is required to substantiate that the copyright holder has not opted out, necessitating active searches for and documentation of relevant opt-outs."). Though the effect of an author's opt-out has been interpreted as prospective, developers must ensure they assess the existence of a reservation at the appropriate point in time (i.e., if there is a delta between assessment of a work's reservation status and use of that work for text and data mining, the author's permission may have changed). *See id.* at 501–02 ("[A] reservation's impact is prospective; if altered subsequently, reproductions already completed remain legal . . . Therefore, opt-outs need only be assessed when initiating new reproductions." (footnote omitted)).

79. *See* DSM Directive, *supra* note 17, at art. 4, ¶ 1.

80. *See id.* (stating that the Article 4 exception only applies if the work in question "has not been expressly reserved by [its] rightholders").

81. *See* DSM Directive, *supra* note 17, at art. 4, ¶ 3 ("The exception . . . shall apply on condition that the use of [works] has not been expressly reserved by their rightholders."); *see* Ilya Ilin & Aleksei Kelli, *Natural Language, Legal Hurdles: Navigating the Complexities in Natural Language Processing Development and Application*, 17 J. U. LAT. 44, 57 (2024) (referring to Article 4's right reservation as an "opt-out right, enabling the exclusion of TDM activities" (footnote omitted)).

any teeth, TDM developers must check for a reservation before using an author's work.[82]

If the DSM Directive's own logic is to be followed, then Article 4's reservation function leads to the same outcome the Directive promised to avoid: increased legal uncertainty.[83] Just as Article 3 was intended to solve for a legal landscape where member states made different legislative decisions regarding copyright exceptions,[84] under Article 4, rightholders will invariably make different choices regarding authorization, creating an EU-wide patchwork of permissions similar to the legislative patchwork pre-Article 3. Where exceptions pre-Article 3 were insufficient because they were "optional and not fully adapted to the use of technologies in scientific research,"[85] rightholder authorization under Article 4 is optional and not fully adapted to the needs of creators or developers.[86] Legal uncertainty pre-Article 3 arose from the fact that "the terms of the licences could exclude text and data mining."[87] Under Article 4, rightholders can similarly exclude text and data mining.[88] Conditions parallel to those that were present and that led to legal uncertainty pre-Article 3 are present under Article 4.[89] Where rightholder authorization is required, then, the legal landscape is definitionally unharmonized and uncertain.

---

82. *See* Lobling, *supra* note 78, at 501 (stating TDM users' responsibility to check for author opt-outs). The EU AI Act confirms, "[w]here the right[] to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works." AI Act, *supra* note 59, at recital 105. Note that the AI Act defines a "provider" as a person or entity "that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark." *Id.* at art. 3, ¶ 3.

83. *See* DSM Directive, *supra* note 17, at recital 11 (stating that Article 3 addresses legal uncertainty); *id.* at recital 18 (stating that Article 4 will "provide for more legal certainty").

84. *See* Giorgos Vrakas, *A literature review of "lawful" text and data mining*, 4 OPEN RSCH. EUR. 1, 7 (2024) ("In a pre-[DSM Directive] era, several Member States had already introduced their own TDM-specific copyright exceptions . . . . As per Geiger et al., 'this regulatory patchwork [. . .] led to a fragmented legal environment in the EU', which ultimately led to the introduction of the [DSM Directive] exceptions." (omission in original) (emphasis omitted) (quoting Geiger et al., *supra* note 67 (manuscript at 24))); *see also* Kanellopoulou-Botti et al, *supra* note 76, at 10 ("[M]ost exceptions or limitations to copyright in the EU legal framework are non-mandatory, [and] they are not implemented the same in EU Members' legal systems . . . . Therefore, there has been in Europe legal uncertainty concerning TDM . . . which . . . the DSM [Directive] aims to alleviate.").

85. DSM Directive, *supra* note 17, at recital 10.

86. *See infra* notes 95–98 and accompanying text (elaborating on the potential inaccessibility and ineffectiveness of the reservation process for rightholders); *see infra* notes 99–100 and accompanying text (explaining how the lack of clarity regarding rightholder reservation mechanics may create confusion or accessibility issues for developers).

87. DSM Directive, *supra* note 17, at recital 10.

88. *See id.* at art. 4, ¶ 3 (allowing rightholders to "reserve" use of their works for text and data mining).

89. *See supra* notes 84–88 and accompanying text (describing the parallels between the legal uncertainty the DSM Directive purports to address via Article 3 and the legal uncertainty it causes via Article 4).

Moreover, the Directive gives little guidance on the mechanics of a rightholder reservation, further exacerbating this legal uncertainty and making the reservation carve-out impracticable.[90] What guidance the Directive does provide is in Article 4(3) and Recital 18. Article 4(3) specifies that rightholders must reserve use of their works "in an appropriate manner, such as machine-readable means in the case of content made publicly available online."[91] Recital 18 further states that these machine-readable means may "includ[e] metadata and terms and conditions of a website or a service" — though "[i]n other cases, it can be appropriate to reserve the rights by other means, such as contractual agreements or a unilateral declaration."[92] Within this menu of options, there is no mention of suggested phrasing. An author wishing to reserve their work's full copyright protection may not know the best way to communicate their reservation, including the wording that is most effective or easily interpretable. An improperly communicated reservation may increase the likelihood that a TDM developer mistakenly trains their model using content that was intended to be reserved by the owner, or that a TDM developer does not train on unreserved content it construes as reserved.

Though the latter may seem like a particularly unlikely scenario, Creative Commons did not seem to think so. In 2021, it published a four-page statement using textual and purpose-based arguments, along with a nineteen-step flow chart, to show that the terms of Creative Commons licenses "cannot be construed or interpreted as a reservation of a right in the context of Article 4 of the [DSM Directive]."[93] This multi-page explanation disclaiming any reservation of rights demonstrates that the terms of a license may ambiguously communicate the reservation status of a work under Article 4. The consequences of getting this wrong could be costly for TDM developers,[94] and the DSM Directive's

---

90. See Gina Maria Ziaja, *The Text and Data Mining Opt-out in Article 4(3) CDSMD: Adequate Veto Right for Rightholders or a Suffocating Blanket for European Artificial Intelligence Innovations?*, 19 J. INTELL. PROP. L. & PRAC. 453, 456 (2024) ("[The DSM Directive] only contains minor indications as to how the reservation of rights should be exercised . . . it appears difficult to ensure that no legally effective reservation has been made.").

91. DSM Directive, *supra* note 17, at art. 4, ¶ 3.

92. *Id.* at recital 18.

93. Ana Lazarova, Thomas Margoni, Ariadna Matas, Sarah Pearson, Julia Reda, Brigitte Vézina et al., *Creative Commons Statement on the Opt-Out Exception Regime / Rights Reservation Regime for Text and Data Mining under Article 4 of the EU Directive on Copyright in the Digital Single Market*, CREATIVE COMMONS (Dec. 17, 2021), https://creativecommons.org/wp-content/uploads/2021/12/CC-Statement-on-the-TDM-Exception-Art-4-DSM-Final.pdf [https://perma.cc/3VWT-GXRN] (clarifying in case of confusion that the language of Creative Commons licenses "do[es] not override exceptions and limitations" and that any conflicting interpretation "runs contrary to the overall design and purpose of the licenses").

94. *See* Theresa M. Weisenberger, Chad A. Rutkowski, Diana C. Milton, Harrison A. Enright & Jiwon Kim, *Case Tracker: Artificial Intelligence, Copyrights and Case Actions*, BAKERHOSTETLER, https://www.bakerlaw.com/services/artificial-intelligence-ai/case-tracker-artificial-intelligence-copyrights-and-class-actions/ [https://perma.cc/XL5P-R59L] ("The rise of advanced generative AI has spawned a flurry of copyright litigations.").

lack of guidance is only one of many risk factors. The DSM Directive implicates additional concerns about accessibility, workability, enforcement, and remediation.

First, enacting the machine-readable reservation options listed above involves know-how and resources that are inaccessible to many creators. Difficulty understanding how to apply a technical law may contribute to legal uncertainty.[95] Take, for example, the process of embedding metadata to signal copyright protection. Not only does this involve technological knowledge to implement in the first place, it also "is not permanent and can be automatically stripped out."[96] Machine-readable reservations can thus fail if an author's work is extracted from a source other than the author herself.[97] Despite best efforts to appropriately reserve one's work, this means it is nonetheless possible for that work to be reproduced elsewhere on the Internet, unattached from its reservation.[98]

These implementation difficulties may lead to confusion for TDM developers regarding where they have or have not received proper authorization for their use of copyrighted works.[99] Implementation difficulties may also create perverse incentives: works may become less accessible to TDM developers if creators take technical measures (such as hiding their content behind a paywall) to ensure it is not used without their authorization.[100]

As for enforcement, it appears the onus is on rightholders to protect their work from unauthorized uses. Recital 18 states that "[r]ightholders should be able to apply measures to ensure that their reservations in this

---

95. *See* Peter H. Schuck, *Legal Complexity: Some Causes, Consequences, and Cures*, 42 DUKE L.J. 1, 18 (1992) (stating that legal uncertainty is dependent on legal complexity); *id.* at 3 (clarifying that legal complexity is in turn comprised of a mix of four factors: "density, technicality, differentiation, and indeterminacy"); *id.* at 5 ("Complexity is multi-dimensional, and its dimensions cannot easily be measured, much less weighted."); *id.* at 4 (clarifying that "*technical* rules require special sophistication or expertise on the part of those who wish to understand and apply them").

96. *See* ASS'N OF MED. ILLUSTRATORS, METADATA–THE KEY TO PROTECTING COPYRIGHT IN DIGITAL IMAGERY 1 (2016), https://www.ami.org/images/stories/documents/AMI_Metadata_Whitepaper_2016.pdf [https://perma.cc/AC7D-U7Y6] (explaining that metadata can be stripped "by 'save for web' functions and by posting on file sharing and social media platforms" and that "[i]t can also be removed and/or replaced by publishers and other end users of images").

97. *See* Ziaja, *supra* note 90, at 456 ("It is often the case that AI systems are trained with pre-prepared datasets that are not accompanied by the rightholders' reservation. However, reservations may still exist regarding the lawfully accessible materials obtained from the original sources . . . [T]he prospect of an effective reservation being made somewhere other than where the material used was acquired remains." (footnote omitted)).

98. *Id.*

99. *See id.* at 456 (discussing the possibility that reservations become detached from the work to which they apply as "it appears difficult to ensure that no legally effective reservation has been made when utilizing large amounts of training materials").

100. *See* Hamann, *supra* note 24, at 118–19 (describing various "technical protection measures" rightholders may employ "to deter crawlers," including paywalls and CAPTCHAs, and other "anti-TDM practices" that disrupt the functioning of TDM techniques).

regard are respected."[101] But it is unclear what a rightholder can do to ensure respect, especially considering the opacity of TDM uses. How is a rightholder to know whether their data is extracted for purposes of scientific research — a use from which they cannot opt out?[102] How are they to know if the reproduction is "transient" and thus excepted under Article 5(1) of the InfoSoc Directive?[103] This system puts a heavy burden on rightholders to ensure that their work is used in the way they intend, with no accompanying guidance regarding how they can supervise these uses.[104]

Even if an author somehow knows their work is being used for unauthorized TDM in violation of their reservation, it is unclear whether there is a feasible remedy. For one, though research is ongoing,[105] selective forgetting remains a "formidable technical challenge" for general-purpose AI ("GAI") models[106] — in other words, artificial intelligence may not be capable of "forgetting" the same way the concept is conceived by humans.[107] So, while other infringements of copyright may be resolved via removal of content, and while an author can injunct *further* extraction of their copyrighted work, once that work is used in the training of an LLM, there may not be a favorable or fair solution for either party.[108]

---

101. DSM Directive, *supra* note 17, at recital 18.

102. *Id.* at art. 3 (excepting "[t]ext and data mining for the purposes of scientific research" from the right of reproduction without the inclusion of a right reservation mechanism for rightholders); *id.* at recital 18 (describing this exception as "mandatory" for rightholders).

103. *See* InfoSoc Directive, *supra* note 17.

104. *See* DSM Directive, *supra* note 17, at recital 18 (prescribing only that "[r]ightholders should be able to apply measures to ensure that their [right reservations] are respected" but including no mention of how rightholders might identify illicit uses).

105. A technique has emerged that appeared effective at erasing memories of the "Harry Potter" novel series from an LLM. *See* Ronen Eldan & Mark Russinovich, Who's Harry Potter? Approximate Unlearning in LLMs 14 (Oct. 4, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2310.02238 [https://perma.cc/2DUP-QYSA]. However, the "density of unique terms or phrases" in the Harry Potter universe may have uniquely contributed to the efficacy of this technique, which may be challenging to extend to non-fiction material, textbooks, or other types of content — much of which may be protected by copyright. *Id.*

106. Cheng-chi Chang, *When AI Remembers Too Much: Reinventing the Right to be Forgotten for the Generative Age*, 19 WASH. J.L., TECH. & ARTS 24, 36 (2024) (explaining that the technical difficulty of selective forgetting is attributable to "[t]he scale of the datasets used to train such models . . . and the interconnectedness of this information within the neural network").

107. *See* Eduard Fosch Villaronga, Peter Kieseberg & Tiffany Li, *Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten*, 34 COMPUT. L. & SEC. REV. 304, 308 (2018) ("AI does not 'forget' data in the way that humans do.").

108. Model deletion is one remedy litigants have suggested in United States courts, including in *The New York Times*' lawsuit against OpenAI for alleged copyright infringement in the training of its large language models. *See* Daniel Wilf-Townsend, *The Deletion Remedy*, 103 N.C.L. REV. (forthcoming 2025) (manuscript at 18). Beyond the more traditional method of data deletion where data is obtained or used illicitly, this method "seeks the deletion of algorithms or models derived 'in whole or in part' using that data." *Id.* at 25. This can be a grossly

III. ARTICLE 4 VIOLATES THE BERNE CONVENTION

Legal uncertainty for rightholders and TDM developers is one problematic effect of Article 4; legal violation is another. Article 4 violates the Berne Convention on two grounds: (1) it does not pass Article 9(2)'s three-step test for exceptions to the right of reproduction,[109] and (2) it violates Article 5(2)'s prohibition of formalities.[110]

A. Article 4 of the DSM Directive Violates Article 9(2) of the Berne Convention

The three-step test in Article 9(2) of the Berne Convention governs whether exceptions to the exclusive right of reproduction are permissible.[111] The limitation or exception: (1) must be confined to "certain special cases," (2) cannot "conflict with a normal exploitation of the work," and (3) cannot "unreasonably prejudice the legitimate interests of the right holder."[112] Because Article 4 excepts "reproductions . . . for the purposes of text and data mining"[113] it is subject to Article 9(2)'s three-step test.

In June 2000, a WTO panel engaged in the first official analysis of the application of Article 9(2)'s three-step test.[114] The dispute in question was over the United States's exemptions under Section 110(5) of the U.S. Copyright Act.[115] These exemptions allowed certain qualifying eating, drinking, and retail establishments to take part in public performances of copyright-protected works.[116] For example, one exemption allowed these establishments to play copyrighted music without requiring authorization or fee payment if the size of the

---

disproportionate remedy, especially in the context of GAI models where vast quantities of data may be used, only some of which may have been illicit. *See id.* at 26.

109. *See infra* Section III.A.

110. *See infra* Section III.B.

111. *See* Ginsburg, *supra* note 54, at 3 ("Berne art. 9.2/TRIPs art. 13 impose the 'three-step test' to evaluate the legitimacy of exceptions and limitations on copyright . . . .").

112. DS160 Panel Report, *supra* note 50, at § 6, ¶ 6.31 (quoting TRIPS Agreement, *supra* note 34, at § 1.1).

113. DSM Directive, *supra* note 17, at art. 4, ¶ 1.

114. *See* Ginsburg, *supra* note 54, at 3 ("The WTO panel decision marks the first time an international adjudicative body has interpreted either art. 13 of TRIPs, or art. 9.2 of the Berne Convention . . . .").

115. *See id.* at 4; *see also* DS160 Panel Report, *supra* note 50, at § 1, ¶ 1.1.

116. *See* Copyright Act of 1976, 17 U.S.C. § 110 (creating an exception for public performances and displays by establishments using a "single receiving apparatus of a kind commonly used in private homes" and an exception for the public performance and display of nondramatic musical works in retail establishments under 2,000 square feet and restaurants under 3,750 square feet, with larger establishments remaining eligible for the exception subject to a limit on the number of devices used to communicate the performance or display).

establishment was less than a specified threshold.[117] The European Communities alleged that these exemptions violated multiple articles of the Berne Convention, including the exclusive right of public performance under Berne Article 11(1).[118] Under the first step of the Article 9(2) analysis, the WTO panel found that one of the exemptions in question did not qualify as a special case.[119] For that exemption, the panel resolved that they need not continue the analysis to "conclude that the business exemption does not satisfy the requirements of [the three-step test], given that its three conditions are cumulative."[120] According to the panel, "[f]ailure to comply with any of the three conditions results in the . . . exception being disallowed."[121] That pattern is echoed here: Article 4 of the DSM Directive fails at the first prong because it is not a special case.

The first factor, requiring exceptions to be confined to "certain special cases,"[122] was read by the WTO panel to require that the limitation or exception is both "clearly defined" and "narrow in its scope and reach."[123] Here, it is clearly defined.[124] There is little question regarding the subject matter covered by Article 4,[125] and relevant terms are defined elsewhere in the directive. Though the rightholder reservation

---

117. *Dispute Settlement Summary — Section 110(5) of US Copyright Act*, WORLD TRADE ORG., https://www.wto.org/english/tratop_e/dispu_e/cases_e/ds160_e.htm [https://perma.cc/ JB5F-TGBJ] ("The so-called 'business' exemption, provided for in sub-paragraph (B) of Section 110(5), essentially allows the amplification of music broadcasts, without an authorization and a payment of a fee, by food service and drinking establishments and by retail establishments, provided that their size does not exceed a certain square footage limit.").

118. *See* Annex I to Dispute Settlement 160: United States — Section 110(5) of the US Copyright Act, *Request for the Establishment of a Panel by the European Communities and their Member States*, at 71, WTO Doc. WT/DS160/5 (Apr. 16, 1999), https://www.wto.org/english/news_e/news00_e/1234db.pdf [https://perma.cc/T9FD-BNPM] ("[Berne] grants the authors of musical works the exclusive right of authorising the public performance of their works . . . . As a consequence of the above, Section 110(5) of the United States Copyright Act appears to be inconsistent with the United States' obligations under the TRIPS Agreement . . . .").

119. World Trade Org., Dispute Settlement Summary — Section 110(5) Copyright Act, WORLD TRADE ORG., https://www.wto.org/english/tratop_e/dispu_e/cases_e/1pagesum_e /ds160sum_e.pdf [https://perma.cc/HYH5-HZXJ] ("[T]he [business] exemption did not qualify as a 'certain special case' . . . .").

120. DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(iv), ¶ 6.160. Note that despite this conclusion, the panel chose to continue the analysis through all three steps of the test, citing potential helpfulness to the dispute settlement body in reaching a decision and general standards for dispute resolution panels as their reasons for continuing. *See id.*

121. *Id.* at § VI(D)(1)(b)(i), ¶ 6.97.

122. *Id.*

123. *Id.* at § VI(D)(2)(b)(ii), ¶ 6.112.

124. The panel provides the following guidance as to the meaning of a "clearly defined" exception: "[T]here is no need to identify explicitly each and every possible situation to which the exception could apply, provided that the scope of the exception is known and particularised. This guarantees a sufficient degree of legal certainty." *Id.* at § VI(D)(2)(b)(ii), ¶ 6.108.

125. DSM Directive, *supra* note 17, at art. 4, ¶ 1 (defining the subject matter covered by Article 4 as "lawfully accessible works and other subject matter for the purposes of text and data mining.").

feature has potentially confusing implications, this prong has not been interpreted as a test of the legitimacy of an exception's policy objective.[126] These implications are thus irrelevant to the exception's analysis under the first step of Article 9(2).[127]

However, though it may be clearly defined, Article 4's exception is not narrow. In its analysis, the WTO panel determined the narrowness of each exception's "scope and reach" by considering "what percentage of eating and drinking establishments and retail establishments may benefit from [or take advantage of]" the exemption.[128] Whereas the beneficiaries of the exception considered by the WTO panel were restaurant and retail establishments,[129] under Article 4 it is those who take part in TDM — TDM developers — that are exempted from copyright infringement.[130] Therefore, the analogous question here is as follows: what percentage of TDM developers may benefit from or take advantage of Article 4's exception?

The answer to this question is simple: one hundred percent of TDM developers may benefit from Article 4's exception. According to Recital 18 of the DSM Directive, "text and data mining techniques are widely used both by private and public entities to analyse large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies."[131] This includes the development of technologies such as OpenAI's GPT-3, which itself used "over 8 million website pages as training data."[132] Whereas Article 3's exception for "text and data mining for the purposes of scientific research"[133] is cabined to specific types of uses, and therefore benefits specific types of TDM developers, Article 4 broadly excepts uses of all kinds of

---

126. *See* DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(ii), ¶ 6.112 (stating with regard to the relevance of the exception's public policy purpose, "a limitation or exception may be compatible with the first condition even if it pursues a special purpose whose underlying legitimacy in a normative sense cannot be discerned. The wording of [the] first condition does not imply passing a judgment on the legitimacy of the exceptions in dispute").

127. *Id.*

128. *Id.* at ¶ 6.112–13 (noting that the report does not contain a justification for using this metric to determine the exception's narrowness).

129. *Id.* at § II, ¶ 2.10 (identifying the beneficiaries of the exception ultimately found in violation of the Berne Convention as "retail establishments" and "food service and drinking establishments").

130. *See* DSM Directive, *supra* note 17, at art. 4 (excepting "reproductions and extractions . . . for the purpose of text and data mining"). This Note defines "TDM developers" as "those engaged in text and data mining." *See supra* Part I.

131. DSM Directive, *supra* note 17, at recital 18.

132. Xuanhe Zhou, Zhaoyan Sun & Guoliang Li, *DB-GPT: Large Language Model Meets Database*, 9 DATA SCI. & ENG'G 102, 102 (2024).

133. DSM Directive, *supra* note 17, at art. 3.

copyrighted works for all kinds of purposes by all kinds of TDM developers.[134]

The WTO panel decided that an exception covering 27–72% of eating, drinking, or retail establishments was too broad to qualify as a special case,[135] but that exceptions covering 13.5–18% of such establishments were sufficiently limited and therefore did qualify.[136] With one hundred percent of TDM developers as potential beneficiaries,[137] Article 4 falls well above the former range. Whereas the exception that the WTO panel analyzed benefitted establishments only when they fell within specific parameters, such as square footage requirements,[138] no equivalent qualification is provided under Article 4.[139] All TDM uses and users are excepted.[140] Therefore, Article 4 fails to qualify as a "certain special case."[141] With the failure of the first condition, the WTO panel instructs that Article 4's exception fails the three-step test, and the analysis need not go further.[142]

Of course, Article 4(3) contains a condition of express reservation that some scholars argue "drastically reduces the number of potential

---

134. Article 4 describes the type of copyrighted work that can be mined under the exception as "lawfully accessible." *Id.* at art. 4. As for purpose, Article 4 only states that reproductions must be made "for the purposes of text and data mining." *Id.* Contrast this to Article 3, which states that the covered text and data mining must be "for the purposes of scientific research." *Id.* at Art. 3. *See* Ted Shapiro & Sunniva Hansson, *The DSM Copyright Directive: EU Copyright Will Indeed Never Be the Same*, 41 EUR. INTELL. PROP. R. 404, 405 (2019) ("Anyone can benefit from the TDM exception: it is not limited to a specific beneficiary."); Eleonora Rosati, *No Step-Free Copyright Exceptions: The Role of the Three-Step in Defining Permitted Uses of Protected Content (Including TDM for AI-Training Purposes)*, 46 EUR. INTELL. PROP. REV. 262, 270 ("Article 3 of the directive provides for an exception . . . that is purpose-limited . . . . Article 4 introduces an exception or limitation that has no restrictions in terms of beneficiaries and purposes of the TDM to be undertaken." (footnote omitted)).

135. *See* DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(ii), ¶ 6.118, 6.133 (finding that the exception benefitting "65.2 per cent of all eating establishments; [] 71.8 per cent of all drinking establishments; [] and 27 per cent of all retail establishments" "does not qualify as a 'certain special case'").

136. *See id.* at § VI(D)(2)(b)(iii), ¶ 6.142–43 (finding that the exception benefitting "16 per cent of all US eating establishments; [] 13.5 per cent of all US drinking establishments; [] and 18 per cent of all US retail establishments" was "limited to a comparably small percentage of all eating, drinking and retail establishments in the United States.") The exception was later found to be "limited in its scope and reach" and therefore "confined to certain special cases." *Id.* at § VI(D)(2)(b)(iii), ¶ 6.159.

137. *See* Shapiro & Hansson, *supra* note 134, at 405 ("Anyone can benefit from the TDM exception: it is not limited to a specific beneficiary."); *supra* notes 128–34 and accompanying text (determining that all TDM developers are potential beneficiaries of the Article 4 exception).

138. *See* 17 U.S.C. § 110(5).

139. *See supra* note 134 and accompanying text.

140. *Id.*

141. *See supra* notes 119–40 and accompanying text.

142. *See* DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(iv), ¶ 6.160 (concluding that failure of the first condition of the three-step test does not require continued analysis of the remaining two steps).

beneficiaries" of the exception.[143] This conclusion, however, is logically incorrect. For the number of potential beneficiaries of the exception to decrease by even one TDM developer, that developer would have to be blocked from taking advantage of every possible piece of copyrighted work. While any one rightholder could plausibly choose to exercise their opt-out right and block every single TDM developer from legally using their work for mining, it is implausible that every single rightholder would do so. As long as at least one rightholder fails to opt out, allowing unfettered TDM access to their work, then every single TDM developer subject to the DSM Directive is a potential beneficiary.[144] While it can reasonably be argued that the opt-out mechanism may reduce the number of copyrighted works that each beneficiary can access, it is illogical to suggest that the mechanism will reduce the number of beneficiaries.[145]

Though Article 4's failure of the three-step test would not be cured by passing steps two and three, if the analysis were to continue, it is plausible that the exception would pass step two. Step two provides that the exception must "[n]ot conflict with a normal exploitation of the work."[146] Uses of one's work for TDM, despite "involv[ing] commercial gain" may not "necessarily conflict[] with a normal exploitation of that work."[147] If TDM is not considered a normal way of "extract[ing] economic value" from a work's copyright, its use would not "deprive [the rightholder] of significant or tangible commercial gains" because the TDM use would not compete with the rightholder's normal uses.[148] As more rightholders choose to license their works to AI companies, however, evidence may be mounting that TDM uses are normal exploitations.[149] Even so, some scholars contend that a reservation system

---

143. Juan-Carlos Fernández-Molina & Fernando Esteban de la Rosa, *Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?*, 24 PORTAL: LIBR. & ACAD. 653, 664 (2024).

144. Note that whether or not potential beneficiaries actually take advantage of an exception has not been held to be relevant to the step one analysis. *See* DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(ii), ¶¶ 6.126, 6.127 ("[W]e do not consider the US calculations of establishments to be deducted from the [beneficiary] estimates as relevant for ascertaining the potential scope of the business exemption in relation to the first condition." (footnote omitted)).

145. *See* Christophe Geiger, Daniel Gervais & Martin Senftleben, *The Three-Step Test Revisited: How to Use the Test's Flexibility in National Copyright Law*, 29 AM. U. INT'L L. REV. 581, 593 (2013) (stating that "number of potential beneficiaries" is the relevant quantitative measure in an analysis of the first step).

146. DS160 Panel Report, *supra* note 50, at § VI(D)(2)(b)(iv), ¶ 6.162.

147. *Id.* at § VI(D)(2) (c)(i), ¶ 6.182.

148. *Id.* at § VI(D)(2) (c)(i), ¶ 6.183.

149. *See* Bill Rosenblatt, *The Media Industry's Race To License Content For AI*, FORBES (July 19, 2024), https://www.forbes.com/sites/billrosenblatt/2024/07/18/the-media-industrys-race-to-license-content-for-ai/ [https://perma.cc/S7T5-4L6D] ("AI content licensing initiatives abound. More and more media companies have reached license agreements with AI companies individually. Several startups have formed to aggregate content into large

may be able to minimize "the potential adverse effect on a work's normal exploitation."[150] Indeed, Article 4 theoretically gives rightholders the ability to reserve their work's copyright, thereby precluding other parties from engaging in conflicting exploitations of their work and preserving their ability to exploit the work themselves.[151] For that reason, the exception may not conflict with a rightholder's normal exploitation, and may pass step two.[152]

The final prong of the three-step test dictates that the exception must "not unreasonably prejudice the legitimate interests of the right holder."[153] This is the case when the exception "causes or has the potential to cause an unreasonable loss of income to the [rightholder]."[154] This is a factual inquiry.[155] Though the DSM Directive benefits from hindsight, it is nevertheless "difficult[] [to] quantify[] the economic value of potential prejudice," and little guidance was provided by the WTO panel regarding the best parameters for this calculation.[156] As with step two, though, it is possible that an exception's "likelihood of satisfying the third step [may be] enhanced if the author's ability to opt out diminished any prejudice otherwise incurred."[157]

B. Article 4 of the DSM Directive Violates Article 5(2) of the Berne Convention

The DSM Directive's Article 4 not only violates the Berne Convention's Article 9(2) requirements for the permissible scope of an exception. It also violates Berne's process requirements. Article 5(2) of the Berne Convention states that "[t]he enjoyment and the exercise of

---

collections for AI platforms to license in one-stop shopping arrangements known in the jargon as blanket licenses."); DS160 Panel Report, *supra* note 50, at § IV(D)2(c)(i), ¶ 6.188 (stating that, while not dispositive, "the extent of exercise or non-exercise of exclusive rights by right holders at a given point in time is of great relevance for assessing what is the normal exploitation with respect to a particular exclusive right in a particular market").

150. M. R. F. Senftleben, *How to Overcome the Normal Exploitation Obstacle: Opt-Out Formalities, Embargo Periods, and the International Three-Step Test*, 1 BERKELEY TECH. L.J. 1, 13 (2014).

151. *See* Hamann, *supra* note 24, at 106–107 (describing the intention of Article 4's right reservation "to nudge parties into bargaining, thereby instrumentalizing unilateral reservations as a conduit to create a (demand-driven) market for TDM licenses"); *see* Ilin & Kelli, *supra* note 81, at 57 ("Essentially, the opt-out right affords rightsholders the opportunity to obtain remuneration for TDM activities.").

152. *See supra*, notes 147–51 and accompanying text (explaining why Article 4 may not conflict with rightholder's normal exploitation of their work).

153. DS160 Panel Report, *supra* note 50, at § IV(D)(1)(b)(i), ¶ 6.31.

154. *Id.* at § IV(D)(1)(d)(i), ¶ 6.229.

155. *See id.* at § IV(D)(1)(d)(i), ¶ 6.236 ("We will consider the information on market conditions provided by the parties taking into account, to the extent feasible, the actual as well as the potential prejudice caused by the exemptions, as a prerequisite for determining whether the extent or degree of prejudice is of an unreasonable level.").

156. *Id.* at § IV(D)(1)(d)(ii), ¶ 6.251.

157. Ginsburg, *supra* note 42, at 761.

[the exclusive rights protected under Berne] shall not be subject to any formality."[158] Though this prohibition against the institution of formalities applies to "foreign authors,"[159] it should be noted that "member states tend not to impose disabilities on their own authors from which they spare foreign authors."[160]

Article 5(2)'s prohibition against formalities has been interpreted to mean that copyright protection under Berne is "automatic"[161] and "vests immediately upon creation."[162] Far from this promised automatic protection, Article 4(3)'s right-reservation regime conditions copyright protection on an author's express positive action,[163] whether that is editing the terms and conditions of a webpage to reserve the right to protect their work from TDM, or embedding metadata to do so.[164] This "express reservation opt-out" requirement is not Berne-compatible because it "condition[s] the 'extent of protection'" of the exclusive right of reproduction promised to creators.[165] Though it is different in form from typical examples of formalities, like notice and registration

---

158. Berne Convention, *supra* note 16, at art. 5, ¶ 2.

159. Ginsburg, *supra* note 42, at 746.

160. *Id.* Furthermore, TDM often involves cross-border interaction such that questions about the copyrightability of works used for TDM may tend to implicate foreign authorship anyway. *See* Liane Colonna, *Opportunities and Challenges to Utilizing Text-Data Mining in Public Libraries: A Need for Legal Research*, 65 SCANDINAVIAN STUD. L. 194 (2018) ("The ability to conduct meaningful research utilizing TDM requires not only being able to access data remotely but also being able to share and further process results within the Digital Single Market and beyond."); *see also* Sean Fiil-Flynn, Brandon Butler, Michael Carroll, Or Cohen-Sasson, Carys Craig & Lucie Guibault et al., *Legal Reform to Enhance Global Text and Data Mining Research*, 378 SCI.: POL'Y F. 951 (2022) ("Although applications of TDM often occur across borders, with researchers, subjects, and materials in more than one country, a patchwork of copyright laws across jurisdictions limits where and how TDM research can occur."). In fact, the DSM Directive places Article 4 under a heading proclaiming to list "measures to adapt exceptions and limitations to the digital and cross-border environment." DSM Directive, *supra* note 17, at title II.

161. *Summary of the Berne Convention for the Protection of Literary and Artistic Works (1886)*, WORLD INTELL. PROP. ORG., https://www.wipo.int/treaties/en/ip/berne/summary_berne.html [https://perma.cc/P6K5-R6J2]; *see* Sam Ricketson and Jane Ginsburg, *The Berne Convention: Historical and Institutional Aspects*, *in* INTERNATIONAL INTELLECTUAL PROPERTY 3, 19 (Daniel J. Gervais ed., 2015) ("[A]uthors' rights in a work vest automatically throughout the Berne Union upon the work's creation (or first publication).").

162. Eric E. Johnson, The Economics and Sociality of Sharing Intellectual Property, 94 B.U. L. REV. 1935, 1944 (2014).

163. *See* Péter Mezei, *A Saviour or a Dead End? Reservation of Rights in the Age of Generative AI*, 46 EUR. INTELL. PROP. REV. 461, 465–66 (2024) ("[R]eservations shall be 'expressed' — that is, rightholders shall openly and expressly claim that they retain the right to TDM over their protected expressions.").

164. *See* DSM Directive, *supra* note 17, at recital 18 ("In the case of content that has been made publicly available online, 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.").

165. Ginsburg, *supra* note 42, at 759 (no citation in original but referencing the Berne Convention); *see* Berne Convention, *supra* note 16, at art. 9.

requirements,[166] Article 4(3) is nonetheless an example of a formality that "come[s] in at . . . the back end" and "shape[s] the scope of protection."[167] And though such express reservation requirements may seem to be allowed under Berne Article 10bis(1), that is not the case: 10bis(1) is "a sui generis provision that . . . does not create a basis for generalization."[168]

Some scholars argue that violation of Article 5(2) may be dependent on the exception's violation of Article 9(2), suggesting that an exception that passes the three-step test "could perhaps be made subject to an express reservation condition" since "the reservation would give the author greater rights than Berne requires."[169] Here, however, Article 4's exception *violates* the three-step test, so the author in fact has fewer rights than the minimum promised by Berne.[170] For an opt-out mechanism to be permissible, the exception to which it applies must be one "which is *already* Berne-compatible."[171] Under the default Article 4 exception, a rightholder is effectively left without copyright protection for works used for TDM, though they may be able to reclaim their Berne-promised rights by opting into the offered right reservation.[172] Forcing a rightholder to affirmatively reserve their copyright on top of an already impermissible exception in effect "impose[s a] condition[] on Berne minimum rights" and thus fails under Article 5(2).[173]

Others may argue that it is legal in the first place to use copyrighted work for TDM, such that Article 4's exception merely codifies the state of copyright law. This, too, is untrue. The DSM Directive's own recitals make clear that copyright protection covering the use of one's work for TDM is an existing right[174] being excepted by Article 4.

---

166. *See* Ginsburg, *supra* note 42, at 747 (contrasting the conditioning of rights "on front-door compliance with notice and registration" with the conditioning of rights on "back-door exercise of opportunities to opt-out").

167. *See id.* at 759.

168. *Id.*

169. *Id.* at 761.

170. *See supra* Section III.A (arguing that Article 4 violates the three-step test).

171. Ginsburg, *supra* note 42, at 760.

172. Article 4(1) provides an unqualified exception "for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining." DSM Directive, *supra* note 17, at art. 4(1). This is qualified by the addition of Article 4(3)'s right reservation condition. *See id.* at art. 4(3) ("The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works . . . has not been expressly reserved.").

173. Ginsburg, *supra* note 42, at 760.

174. *See* DSM Directive, *supra* note 17, at recital 8 ("In certain instances, text and data mining can involve acts protected by copyright, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are normalised in the process of text and data mining.").

IV. ARTICLE 4 SETS A DANGEROUS PRECEDENT

The attempt to normalize broad exceptions that fail Article 9(2)'s special case factor and right reservation requirements that function as Article 5(2) formalities is dangerous if left unchecked. For one, the EU's Artificial Intelligence Act ("AI Act") builds upon this framework and further reinforces it — for example, referencing the DSM Directive and reiterating that "rightsholders may choose to reserve their rights . . . to prevent text and data mining" for purposes other than scientific research.[175] This matters: the AI Act is "the world's first comprehensive AI law,"[176] and other countries are expected to "emulate many aspects of it."[177]

Article 4 also makes clear that EU legislators are prioritizing innovation at the cost of author protections. Though Article 4's right reservation mechanism is rightholders' only avenue through which to maintain copyright protection from TDM, in the five years since the DSM Directive was passed, no standardized author opt-out system has been developed by legislators.[178] As the text of the Directive makes clear, legislators were uncomfortable with the "legal uncertainty"[179] copyright rules created for companies wanting to innovate using TDM. No efforts have been made, however, to address the uncertainty and impossibility that Article 4 creates for copyright holders.[180] In this era of artificial intelligence, the EU's choice in the tradeoff between innovation and protection is clear.

Europe is already seeing lawsuits challenging this kind of prioritization. In early 2023, Getty Images filed suit against Stability AI in the United Kingdom, alleging copyright infringement for "cop[ying] at least 12 million copyrighted images from Getty Images' websites, along with associated text and metadata, in order to train its [GAI image

---

175. AI Act, *supra* note 59, at recital 105.

176. European Parliament, *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARLIAMENT: TOPICS (June 18, 2024, 4:29 PM), https://www.europarl.europa.eu/top ics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence [https://perma.cc/H4UZ-VNLR].

177. Kelvin Chan, *Europe Reaches a Deal on the World's First Comprehensive AI Rules*, ASSOCIATED PRESS (Dec. 8, 2023, 6:19 PM PDT), https://apnews.com/article/ai-act-europe-regulation-59466a4d8fd3597b04542ef25831322c [https://perma.cc/XL5Z-TNGQ] (quoting Professor Anu Bradford).

178. *See* Ziaja, *supra* note 90, at 456 ("[T]here are no generally accepted protocols or standards for the machine-readable expression of opting out.").

179. *See, e.g.*, DSM Directive, *supra* note 17, at recital 8 ("[T]here is widespread acknowledgement that text and data mining can, in particular, benefit the research community and, in so doing, support innovation . . . . However, in the Union, such organisations and institutions are confronted with legal uncertainty as to the extent to which they can perform text and data mining of content."). The Directive later states that this legal uncertainty "should be addressed by providing for a mandatory exception." *Id.* at recital 11.

180. *See supra* note 178 and accompanying text (underscoring the inaction of legislators in clarifying the mechanics of a right reservation under Article 4).

generation] model" — despite the fact that Getty Images' website terms of use "expressly prohibit unauthorized reproduction of content for commercial purposes."[181] If right reservations are ignored or misinterpreted, the EU could begin to see similar lawsuits.

Perhaps even more worrisome, artists have claimed that AI image generators training on their works are doing more than just infringing their copyright through training: they are also allegedly generating images that copy the artists' style.[182] Combined with the complexity and costs of AI "forgetting" and "unlearning,"[183] this may comprise a significant harm for both rightholders and infringing developers. Article 4 leaves the door to such harms wide open. In fact, in a July 2023 letter, 24 U.S.-based artistic groups wrote to the United States government urging it to "use all available means to bring the European Union into compliance with the Berne Convention . . . in connection with the application of Articles 3 and 4 to generative AI," alleging harms such as unauthorized and unremunerated mining of their work to create derivative works.[184] Signatories included the Authors Guild and the National Writers Union.[185] More complaints may follow as GAI models become more advanced and as more creators are affected by TDM developers' use of copyrighted work.

V. SOLUTIONS

When a draft of the EU's AI Act was released to the public, OpenAI was clear on their position: they would try their best to comply.[186] But if they couldn't, they might consider ceasing operations in

---

181. Complaint & Demand for Jury Trial at 3, Getty Images, Inc. v. Stability AI, Inc., No. 1:23-cv-00135 (D. Del. Feb. 3, 2023), https://fingfx.thomsonreuters.com/gfx/le galdocs/byvrlkmwnve/GETTY%20IMAGES%20AI%20LAWSUIT%20complaint.pdf [https://perma.cc/48XV-LL7M].

182. *See* Rachel Metz, *These artists found out their work was used to train AI. Now they're furious*, CNN (Oct. 21, 2022, 9:06 AM EDT), https://edition.cnn.com/2022/10/21/tech/art ists-ai-images/index.html [https://perma.cc/C3PG-EM9A] (detailing artists' dissatisfaction upon learning that an AI model was capable of generating images mimicking their style after training on pictures of their work).

183. *See supra* notes 105–08 and accompanying text (explaining that AI does not "forget" the same way people do, and that unlearning methods may be costly for developers).

184. National Writers Union, American Society for Collective Rights Licensing, Artists Rights Society, American Photographic Artists, Romance Writers of America & National Press Photographers Association et al., *Appeal for Action on Violations of the Berne Convention by the Application to Copying of Creative Works for AI Development of the TDM Exception in Articles 3 and 4 of the 2019 EU Directive on Copyright*, NAT'L WRITERS UNION 1, 1 (July 2023), https://nwu.org/wp-content/uploads/2023/07/creators-coalition-AI-exceptions.pdf [https://perma.cc/FRB9-P43G].

185. *Id.* at 4–5.

186. *See* Reuters, *OpenAI May Leave the EU If Regulations Bite — CEO*, REUTERS (May 24, 2023 5:22 PM EDT), https://www.reuters.com/technology/openai-may-leave-eu-if-regu lations-bite-ceo-2023-05-24/ [https://perma.cc/EQW8-P2B3] ("OpenAI CEO Sam Altman

the EU.[187] OpenAI was not shy about its belief that the proposal was a case of "over-regulating."[188]

Statements such as this one from companies like OpenAI — companies that represent the very innovation the EU is trying to attract through measures such as Article 4[189] — make a complete overhaul of Article 4 unlikely. Requiring TDM developers to actively seek or purchase permission to mine all copyrighted works would be burdensome, and developers in the EU might choose to conduct business elsewhere.[190] Such economic incentives militate against author-favorable changes to Article 4. Indeed, in the case of the AI Act, OpenAI's lobbying to exclude GPT-3 from the highest level of regulation appeared successful.[191]

Despite this, the AI Act may represent progress for rightholders. A realistic solution to some of the harms faced by creators under Article 4 may be balancing lower levels of copyright protection with disclosure and transparency requirements for TDM developers. The AI Act achieves this through Article 53, which requires providers of general purpose AI models to "draw up and make publicly available a sufficiently detailed summary about the content used for training" the model.[192] This may enable authors to better track uses of their work and enforce their rights.[193]

Of course, such a benefit depends on the thoroughness of the summary. Recital 107 expands on Article 53's purpose, stipulating that the summary "should be generally comprehensive" to allow copyright owners and other stakeholders "to exercise and enforce their rights."[194] Despite this acknowledgement of copyright owners' specific interest in

---

said on Wednesday the ChatGPT maker might consider leaving Europe if it could not comply with the upcoming artificial intelligence (AI) regulations by the European Union.").

187. *See id.*

188. *Id.*

189. *See* Communication from the Commission to the European Parliament, *supra* note 21, at 11 ("[T]he EU needs to . . . create the right environment for [AI specialists] to work in the EU and attract more talent from abroad." (emphasis omitted)); DSM Directive, *supra* note 17, at art. 4, ¶¶ 1–3.

190. *See* Pamela Samuelson, *Legally Speaking: The EU's Controversial Digital Single Market Directive*, 61 COMMC'NS ASS'N. COMPUTING MACH. 20, 23 (arguing that without a broad TDM exception, "some EU firms may ship their TDM research offshore to take advantage of less-restrictive TDM rules" and "the EU may suffer from 'brain drain' if its most talented researchers take job opportunities in jurisdictions where TDM is broadly legal").

191 Billy Perrigo, *Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation*, TIME (June 20, 2023, 1:00 AM EDT), https://time.com/6288245/openai-eu-lobbying-ai-act/ [https://perma.cc/T72S-AT9R] ("OpenAI's lobbying effort appears to have been a success: the final draft of the Act approved by lawmakers did not contain wording present in earlier drafts suggesting that general purpose AI systems should be considered high risk.").

192. AI Act, *supra* note 59, at art. 53, ¶ 1(d).

193. *See* Ziaja, *supra* note 90, at 457 ("The publication of a comprehensive list of works used for TDM activities will enable rightholders to be aware that their works have been used to train specific AI systems.").

194. AI Act, *supra* note 59, at recital 107.

identifying uses of their works, issues for authors persist. For one, Recital 107 lists as an example of a sufficiently comprehensive summary one that "list[s] the main data collections or sets that went into training the model" and "provid[es] a narrative explanation about other data sources used."[195] It is easy to see how this can complicate authors' task of determining whether their works were used in training. Even an author with an immaculate recollection of the datasets in which their work is embedded will be unable to identify use of their work if it was (1) used without their authorization in the third-party dataset or (2) recounted in the seemingly less granular narrative component of the summary. For authors, Article 53 is an imperfect solution. For TDM developers, however, the undemanding nature of the summary may strike a convenient balance.

Potentially less convenient for TDM developers, but no less important, is another of Article 53's obligations: enacting a policy "to identify and comply with . . . a reservation of rights expressed pursuant to Article 4(3) of [the DSM Directive]."[196] On one hand, it is very much in TDM developers' interest to avoid mistakenly training models on reserved works and the legal trouble that could ensue.[197] In fact, individual companies are already creating their own opt-out systems. Google has a tool allowing website publishers to opt out of Google's training of AI models,[198] and OpenAI claims to be developing a similar tool allowing creators to tailor the use of their work in machine learning training.[199] But for developers that are not so inventive, the burden of *identifying* reservations may be inefficient and burdensome given there is no standardized protocol for *creating* reservations.[200] The issues

---

195. *Id.*

196. *Id.* at art. 53, ¶ 1(c).

197. *See, e.g.*, Jenny Quang, Note, *Does Training AI Violate Copyright Law?*, 36 BERKELEY TECH. L.J. 1407, 1408 ("In a recent example, an AI-based legal research startup shut down amid financial pressures brought on by a copyright infringement lawsuit. And statutory fines, which range from $200 to $150,000 per work, are unnecessarily crippling, especially when a single machine learning model may be trained using thousands to millions of works." (footnotes omitted)).

198. *See* Emma Roth, *Google Adds a Switch for Publishers to Opt Out of Becoming AI Training Data / Now the Google-Extended Flag in Robots.txt Can Tell Google's Crawlers to Include a Site in Search Without Using It to Train New AI Models Like the Ones Powering Bard*, VERGE (Sept. 28, 2023, 3:31 PM EDT), https://www.thev erge.com/2023/9/28/23894779/google-ai-extended-training-data-toggle-bard-vertex [https://perma.cc/7UQR-NJF9] ("The new tool, called Google-Extended, allows sites to continue to get scraped and indexed by crawlers like the Googlebot while avoiding having their data used to train AI models as they develop over time.").

199. *See* OpenAI, *Our Approach to Data and AI*, OPENAI (May 7, 2024) https://openai.com/index/approach-to-data-and-ai/ [https://perma.cc/FRS3-2KUZ] ("OpenAI is developing Media Manager, a tool that will enable creators and content owners to . . . specify how they want their works to be included or excluded from machine learning research and training.").

200. *See* Ziaja, *supra* note 90, at 456 ("At the moment, there are no generally accepted protocols or standards for the machine-readable expression of opting out.").

associated with this lack of direction are numerous.[201] And while
OpenAI's aspiration to "set a standard across the AI industry"[202] is
heartening, government guidance may make widespread adoption more
likely.[203] An EU-wide approach to creating and identifying right reser-
vations would have benefits for both rightholders and TDM developers:
it would streamline the process of communicating a work's copyright
protection and facilitate compliance with the AI Act's Article 53.[204] It
thus may be among the most realistic solutions to the uncertainty
caused by Article 4(3).

More simply than an entirely new protocol, an amendment to the
DSM Directive could make the menu of recommended machine-read-
able options exhaustive rather than open-ended.[205] Some machine-read-
able means are significantly more challenging than others for TDM
developers to interpret; excluding these options may increase compli-
ance with right reservations (and, consequently, Article 53 of the AI
Act).[206]

Of course, this does not account for every issue. For one, it does
not address the lack of accessibility and technical knowledge that
rightholders may be confronted with when attempting to implement a
machine-readable reservation.[207] Second, it does not solve for the

---

201. *See supra* notes 90–94 and accompanying text (describing the negative effects result-
ing from the DSM Directive's lack of guidance regarding the creation of reservations).

202. *See* OpenAI, *supra* note 199.

203. Paul Keller & Zuzanna Warso, *Defining Best Practices for Opting Out of ML Train-
ing*, 5 OPEN FUTURE POL'Y BRIEF 1, 1–2 (2023), https://openfuture.eu/wp-content/up
loads/2023/09/Best-_practices_for_optout_ML_training.pdf   [https://perma.cc/659J-C87Y]
("[Identifying best practices for the communication of opt-outs under Article 4 of the DSM]
requires the intervention of an actor with sufficient credibility to provide guidance on how to
express machine-readable rights reservations. In the current constellation, the entity best
placed to take on this role is the European Commission.").

204. *Id.* (arguing that the EU's intervention in identifying best practices for opting out of
Art. 4 would provide creators with needed guidance on how to express reservations and TDM
developers with more certainty and incentive to respect those opt-outs); AI Act, *supra* note
59, at art. 53, ¶ 1(c) (requiring AI model providers to create a policy to "identify and comply
with" right reservations).

205. *See* Hamann, *supra* note 24, at 110 ("There is a wide range of potential interpretations
of 'machine-readable' . . . . The range of these potential interpretations has caused great un-
certainty.").

206. *See* Lobling et al., *supra* note 78, at 504 ("Effective opt-out management [for reser-
vations within a website's terms of use] would require advanced NLP methods, which might
still carry high error rates. This could undermine the TDM exception's effectiveness. Opting
out in a website's terms of use would not be appropriate to the automated processes of
TDM."); AI Act, *supra* note 59, at art. 53, ¶ 1(c) (requiring providers of AI models to enact a
policy "to identify and comply with . . . a reservation of rights expressed pursuant to Article
4(3)").

207. *See* Hamann, *supra* note 24, at 107 (citing concerns that reservations will be "'difficult
in practice' because 'very few authors have the requisite skills and knowledge to draft a res-
ervation.'" (footnote omitted)); *id.* at 121 ("[T]he large majority of small creators can barely
keep an eye on the developing landscape of reservation standards, let alone properly imple-
ment the requisite standard(s). Only resourceful repertoire owners have the capacity needed
to understand and implement each of the available standards.").

scenario in which TDM developers reproduce or extract works from a third-party source, as existing machine-readable reservation techniques do not appear to address this issue.[208] For the latter issue, a new EU-wide standard remains a better solution, and OpenAI's Media Manager tool may provide a model when developed.[209] The tool "will require cutting-edge machine learning research" and aspires to "help [OpenAI] identify copyrighted text, images, audio, and video across multiple sources."[210] This technology could enable TDM developers to comply with reservations, even when work has been "quoted, reviewed, re-mixed, reposted and used as inspiration across multiple domains."[211]

VI. CONCLUSION

This Note has made the case against the DSM Directive's compatibility with both the needs of stakeholders and the Berne Convention. Article 4 of the DSM Directive is impermissibly overbroad[212] and violates a core tenet of the Berne Convention: that copyright protection is inherent to the creation of a work.[213] For over a century, this principle has been a cornerstone of the international copyright regime.[214] It has fostered a transnational market for authors, offering broad and automatic protection without the imposition of administrative barriers.[215]

Whether the DSM Directive represents a shift toward a new international philosophy or stands as a stark exception to Berne's longstanding principles remains to be seen. However, if policymakers continue to shape AI legislation without engaging with the challenges posed by Berne's underlying ideologies, the dominant copyright principles of the era risk being chosen by political pressures rather than time-tested deliberation.

Furthermore, as the landscape of AI technologies continues to evolve, additional issues appear likely to emerge. For example, questions abound about the copyright ownership of GAI outputs.[216] The AI

---

208. *See* Ziaja, *supra* note 90, at 456 ("Even if the rightholder opted-out in a machine-readable format, the prospect of an effective reservation being made somewhere other than where the material used was acquired remains." (footnote omitted)).

209. *See* OpenAI, *supra* note 199 ("[M]any creators do not control websites where their content may appear, and content is often quoted, reviewed, remixed, reposted and used as inspiration across multiple domains. We need an efficient, scalable solution . . . We're building Media Manager so content owners can manage how their works are used in AI.").

210. *Id.*

211. *Id.*

212. *See supra* Section III.A.

213. *See supra* Section III.B.

214. *See Berne Convention Guide*, *supra* note 33.

215. *See supra* notes 42–47 and accompanying text.

216. *See* Pamela Samuelson, *Generative AI Meets Copyright*, 381 Sci. 158, 161 (2023) (analyzing the likelihood of success of the argument that GAI models' image outputs

Act's implementation may also affect these questions, as provisions like the input data transparency mandate for high-risk systems take effect. It is yet to be seen whether other countries will follow the EU's lead in aggressive and comprehensive regulation of AI. But until then, copyright owners in the EU are left to wonder if these regulations are coming too early. Creators affected by the DSM Directive's Article 4 have certainly shared their misgivings,[217] and if the AI Act creates a particularly favorable regulatory environment for TDM developers, these effects could compound as companies clamor to enter the EU market.

At the same time, the legal uncertainty resulting from regulations like Article 4(3)'s spotty instruction on right reservation may lead to litigation against TDM developers using works without proper permission, similar to the lawsuits already seen in the United States.[218] In theory, these competing influences may themselves regulate such companies' behavior, encouraging them to implement a reliable way to check for reservations when using copyrighted work; in practice, litigation and the threat thereof may not be sufficient deterrents.[219] This emphasizes the crux of the matter: it is the government's job to protect the rights promised to its citizens. With Article 4's continued existence, unamended as if forgotten, the EU is failing at this job.

---

constitute infringement of image-based training data); *see also* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB 10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 1 (2023), https://crsreports.congress.gov/product/pdf/LSB/LSB10922 [https://perma.cc/WAD8-F54D] (exploring "questions that courts and the U.S. Copyright Office have begun to confront regarding whether generative AI outputs may be copyrighted").

217. *See* Hamann, *supra* note 24, at 107 (citing German journalists' concerns about the impracticality of Art. 4's right reservation). *See generally* Nat'l Writers Union et al., *supra* note 184 (expressing concern about DSM Directive Article 4 and its effects on rightholders).

218. *See, e.g.*, Jennifer Korn, *Getty Images Suing the Makers of Popular AI Art Tool for Allegedly Stealing Photos*, CNN (Jan. 18, 2023, 9:47 AM EST), https://edi tion.cnn.com/2023/01/17/tech/getty-images-stability-ai-lawsuit/index.html [https://perma.cc/MG5N-BPFF] (detailing Getty Images' suit against Stability AI for alleged copyright infringement in training its AI art tool); Plaintiffs' Opposition to Defendant Midjourney, Inc.'s Motion to Dismiss Plaintiffs' Complaint and to Strike Class Claims at 1, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. Oct. 30, 2023) (referring to the plaintiff's allegations that AI service Midjourney "directly copied Plaintiffs' images to train [its] AI image generators").

219. *See* Wilf-Townsend, *supra* note 108 (manuscript at 20–21) ("[T]here is no particular reason to think that actual damages or statutory damages would approximate [comparable deterrence to model deletion]. And second, even where damages are set high enough from a deterrence perspective . . . if there is continued value to a company to use the model . . . the company will likely continue to use it.").