

# **AI TRAINING THROUGH COPYRIGHTED WORKS AS INFRINGEMENT: PERSPECTIVES UNDER THE BERNE THREE-STEP TEST AND THE PANE FAIR USE TEST AND PLAUSIBLE SOLUTIONS**

**RUBÉN CALDERÓN**

## **INTRODUCTION**

The human brain is one of the most fascinating components of our current existence. For millennia, it has been capable of different cognitive processes that have become second nature to social life, such as obtaining, processing, and storing information, or learning from experience. Today, the marvels of human creation have gone far beyond creative works and inventions being able to replicate cognitive operations in complex computational systems that many fear will take over several human positions and responsibilities. This includes data processing, decision-making, and even the sacred act of intellectual creation. That is the crux of the generative artificial intelligence (“AI”) debate.

Still, for those applications to be possible, enablement is required. Similar to the human brain, which requires acquiring information through interactions with the environment, available sources, and communication with others, AI needs to be fed with information to learn. To understand the data, it needs to read it; to create, it needs prior knowledge; to write, it needs to know a language comprised of something other than the 1s and 0s of its design. In broad strokes, this enablement process is known as AI training.

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* **85**

However, in practice, AI training has caused several issues and has posed many questions. For example, there is the age-old question: “Who watches the watchmen?” Authors around the globe are distressed over the capabilities of AI, particularly regarding the source of information used to train these applications. This issue is aggravated by the many unresolved political, legal, and corporate challenges that must be addressed before reasonable practices and policies to “watch over” AI companies and their products can be applicable and standardized around the globe.

As previously hinted, the essential question this essay aims to tackle is whether AI training constitutes copyright infringement or if, in the current state of affairs, this process could be covered by a copyright exception. To do so, this essay will provide a brief technical introduction to the process of AI training. This essay will then address the arguments proposed by both camps in light of recent events, such as the Authors Guild of America class action suit against OpenAI and the European Writers’ Council’s declarations and demands to amend European copyright law.

These clashing perspectives will elucidate the proposed purpose of this essay: applying the Berne three-step test and the U.S. fair use test to predict the likely outcome of this situation in both U.S. courts and European legislation, while considering the subtle rational differences between each test. This will aid in identifying challenges and proposing likely solutions. Nonetheless, it must be noted that this is a weighing exercise that involves the clash that has accompanied IP law since its inception: the erroneously created dilemma between the progress of science and useful arts and the rewards authors should earn for their creations. Thus, the solutions proposed herein aim to introduce a comprehensive framework that addresses legal certainty, policy-making, and licensing mechanisms.

## **THE AI TRAINING PROCESS: WHERE IS THE INFORMATION COMING FROM?**

The world is full of information that the human brain learns from. Every book read, class attended, and conversation had provides inputs that enable cognitive processes, such as language learning. This is how the speaker at a conference learned to speak or the reader of this article learned to read. The information stored in the brain is digested and assembled through context and definitions that are symbolized by a particular linguistic system. Artificial Intelligence emulates this process within the capabilities of its application programming interface (“API”) through natural language processing. This is a traditional AI application that allows communication with the system in a particular language. For example, to produce an English text, the AI needs to learn how to communicate in English. To do this, AI operators must feed it different data that will build a large language model (“LLM”).

These models are defined as a type of AI model that can process and generate natural language text. LLMs are typically trained with massive amounts of text and deep learning techniques to learn the patterns and structures of language. Following the significant advancements AI made with the introduction of GPT-3, the use of the functions of LLMs has become increasingly specialized. This specialization has led to what is now referred to as the natural language processing revolution, thanks to the massive scale of LLMs.

However, the scalability of the model and its ability to identify linguistic patterns and structures depend directly on the quantity and quality of the data provided. The provision of said data that is accompanied by programmatic machine learning functions that allow it to carry out specific operations (e.g., read, identify, synthesize, and even create) is defined as Artificial Intelligence training. The crux of the

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* **87**

infringement matter ultimately lies in the data that feeds this training.

AI will not be able to build its LLM and start its machine learning process without its operator or programmer providing it with various packages of data, called data sets. In science fiction films, if the central concern of the protagonists is when Asimov's laws will fail and AI will rebel against its creators, authors and copyright holders beg the following questions: 1) Does AI have indiscriminate access to works protected by copyright law? 2) What can be done with those works through the trained systems? These questions require analyzing the content of the data sets used in training.

Although each AI system and company has its own data sets, these can, for copyright purposes, be categorized into three large groups: Simple data, works in the public domain, and works protected by copyright law. Many times, developers and operators are unaware of the presence of protected works in their data sets. For instance, piracy is directly linked to AI training, since illicit hubs like GenLib and Z-Library are common sources for obtaining data sets. Though the existence of these portals containing billions of gigabytes of illicitly reproduced and distributed works is a problem in itself, the possible infringement analyzed in this essay will be limited to the claimed unauthorized use of copyrighted works in AI training.

## **IS IT INFRINGEMENT? REVIEWING COPYRIGHT AND AI**

Through Copyright, the work produced by an author obtains exclusive rights over their creation. These rights include, among other benefits, the exclusive right to authorize the reproduction, transformation, distribution, and public communication of the work. This is the reward the copyright system gives authors: the power to solely exploit their work as they see fit. For example, creators may license

their works for translation into other languages or for adaptation into other mediums.

During the training process, reproduction rights, public communication rights, and transformation rights are likely to be infringed by AI companies. Nonetheless, it must be noted that the liability of third parties, such as AI users, will not be factored in. Instead, for this study, the implications of the technological practices employed by AI companies and the service provided will be the exclusive focus.

## **REPRODUCTION RIGHTS**

As mentioned above, the AI learning and training process requires the AI to be fed with almost endless batches of data and information. The aforementioned controversy raised by the use of files from GenLib and Z-Library occurs because batches of information called Books1, Books2, and Books3, each protected by copyright, are found in several AI training programs, such as OpenAI's ChatGPT.

Regardless of the link between these data sets and the acts of piracy perpetrated by the administrators of these clandestine digital libraries, the fact remains that the introduction of these works in the LLM of the AI program represents, *prima facie*, an infringement of the right of reproduction. In this computational process, the downloaded data is essentially copied and archived on hard drives, cloud storage, or other data repositories. As it is, AI operators plausibly create at least one copy of the work. The holding of the United States District Court for the Southern District of New York, quoted here in reference, provides a relevant analysis of reproduction in digital environments: "Simply put, it is the creation of a new material object and not an additional material object that defines the reproduction right. The dictionary defines 'reproduction' to

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* 89

mean, *inter alia*, ‘to produce again’ or ‘to cause to exist again or anew.’”

The reproductions that occur during AI training can be summarized in a few steps. First, the AI provider downloads the data set from the net and stores it in its systems. This is Copy A. Second, the AI provider uploads the data to their model, requiring storage of copies of the files, copyrighted or otherwise. This is Copy B, and it will likely be reproduced *ad infinitum* in cloud storage and servers since there is no certainty as to how many copies are required or created during the training process. Under current regulation, this directly infringes the reproduction right.

## **PUBLIC COMMUNICATION RIGHTS**

The creation of the digital market posed an interesting question: Is e-commerce and the market for digital works a form of distribution or public communication? Luckily, international legislation included a very clear solution in the WIPO Copyright Treaty of 1996 through the agreed statement concerning articles 6 and 7. The treaty clearly states: “[T]he expressions ‘copies’ and ‘original and copies’, being subject to the right of distribution and the right of rental under the said Articles, refer exclusively to fixed copies that can be put into circulation as tangible objects.” Therefore, it is clear that any act performed by a digital service will constitute an act of public communication. This includes, among other things, the download of ebooks and other files.

As previously stated, there is certainly at least one copy of the copyrighted work being used for training within the AI program. Depending on the prompt provided to the program and its API configuration, it could perform certain actions that violate the public communication right. For instance, highly trained AI can give the user detailed

summaries or even fragments of the copyrighted work ingested through its data set. This is possible because the illicit reproduction exists within the AI’s “brain.” This copy enables infringing public communication acts, as long as companies do not enforce content locks. This is because “[a]n online platform may be equipped with AI functionalities that select what content is accessible to the public.”

### **TRANSFORMATION RIGHTS**

Regardless of the authorship debate, infringement must be evaluated under the premise that generative AI “creates,” and thus, exercises creativity. Fully generative systems are capable of creating artistic works on their own. They can do so in all artistic fields. Therefore, infringement of the exclusive transformation right is plausible. When AI is prompted to complete actions such as outlining a sequel, translating a poem, or animating a book, it is indeed creating a derivative work. Recent publications have noted that “the very process of AI-based machine learning, which leads to those productions, is arguably a form of derivation.” Like the public communication right, training enables different infringing situations.

It must be noted that the machine learning process, combined with text and data mining applications, allows the AI to understand through language, patterns, and resources that characterize certain copyrighted works. For instance, text-generating AI infringes the transformation right when translating or adapting literary works. The same can be said when image-generating AI adapts copyrighted works to other forms of expression. Given the potential growth of AI applications, the capabilities of AI to generate unauthorized derivative works are currently uncertain.

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* 91

The Author's Response: Authors Guild of America's  
Class Action, New York Times' Lawsuit, and European  
Writers' Council Statements

Reviewing the possible infringed rights allows for understanding the context and reasoning behind three recent relevant events: two lawsuits against OpenAI and statements published by the European Writers' Council. This section gives a summary and relevant analysis of these events. However, the conclusions to be drawn from each lawsuit and the demanded policy reforms in Europe will be looked into in its dedicated section as this will provide the necessary context to apply the three-step test and the PANE fair use test.

**AUTHORS GUILD V. OPENAI INC.**

In September 2023, the Authors Guild, on behalf of the collective of American literary authors and particularly well-known authors such as George R.R. Martin and John Grisham, filed a class action lawsuit against OpenAI and its affiliates in New York. In the lawsuit, the authors alleged infringement of their reproduction and transformation rights. More specifically, by entering prompts requesting summaries, possible sequels, or adaptations of their works, the ChatGPT output generated summaries and accurate transcriptions of the works.

The central argument of the Authors' claim was that "ChatGPT could not have generated the results described above if OpenAI's LLMs had not ingested and been "trained" on the [Authors'] Infringed Works." The crux of this problem lies in the existence of copies of the illicitly obtained works, the reproduction of which was necessary for the training of the model. To date, OpenAI has not made its defenses to the authors' claims public, unlike the case brought by the New York Times.



However, recent press reports that have investigated the extrajudicial evolution of the case indicate that the Authors Guild is exploring the possibility of implementing a blanket licensing system, whereby the companies ingesting the protected works would remunerate the authors. Although publicly available information is scarce, this possibility is one of the proposals to be studied below. Whatever the outcome, judicial or extrajudicial, this class action has the potential to set trends in the literary industry and its treatment of AI training.

### **NEW YORK TIMES CO. v. OPENAI INC.**

On December 27, 2023, the New York Times (“NYT”) filed a lawsuit against OpenAI and its subsidiaries alleging infringement of its copyrights. The NYT claimed that the process of training the “memorization” function induces ChatGPT to reproduce word-for-word articles from the American newspaper. In the lawsuit, memorization is defined as the situation where “[G]iven the right prompt, they will repeat large portions of materials they were trained on. This phenomenon shows that LLM parameters encode retrievable copies of many of those training works.”

In turn, the NYT argued that OpenAI is a free-rider on the newspaper’s investment in research, writing, and technology, as they profit from its content without offering compensation. The NYT pointed out that this free riding of copyrighted journalistic works is probable because in the process of data mining OpenAI’s web-based ChatGPT training, [www.nytimes.com](http://www.nytimes.com) was the most used proprietary source in the AI company’s data sets.

### **EUROPEAN WRITERS’ COUNCIL STATEMENT**

When European Directive 2019/790 entered into force, a newly included copyright exception caused

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* 93

controversy among authors and proprietors. This is none other than the text and data mining (“TDM”) exception contained in Articles 3 and 4 of the Directive. The presence of this new exception to copyright in the European framework is, as is clear from the recitals of the Directive, an exercise of balancing the interests of rightsholders against technological and scientific progress. In the words of the European Parliament, “Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible. . . . [T]here is widespread acknowledgment that text and data mining can, in particular, benefit the research community and, in so doing, support innovation.”

Although well-intentioned, the practical effects of this exception have received fair criticism. For instance, compliance with all the conditions of Article 4 for the exception to apply is, to date, difficult to verify for those wishing to engage in mining. Similarly, it is uncertain as to the appropriate means by which copyright owners may exercise their right to opt-out or the extent to which lawful obtaining of the work is a requirement. It is also uncertain whether, for example, works deposited and preserved in digital libraries are covered by the exception. Similarly:

[S]ince it is possible to exclude the possibility of carrying out TDM when not all the requirements set forth in Article 3 are met, any legal operator will require prior analysis of the legality of a possible agreement and its extension, conditions of use or self-declaration of exclusion of TDM activities.

A notable development in the opposition to this and future upcoming changes to the treatment of AI in the European Union is the response of the European Writers’ Council (“EWC”). The EWC demanded the implementation of the Authorization, Remuneration, and Transparency (“ART”) principle stating that “the Text and Data Mining

Exception(s) (§3, §4, CDSM Directive 2019/790 (EU)) does NOT cover machine learning for especially generative AI and large foundation models.” Further, European authors demanded that the reform considers the Berne Convention’s three-step rule, essential for the limitation of authors’ rights, by requiring that reproductions made to execute data mining operations only be applied in special cases that do not conflict with the normal exploitation of works and do not unreasonably prejudice the legitimate interests of rightsholders.

### **IS IT NOT INFRINGEMENT? OPEN AI’S & CO. FAIR USE CLAIM**

A few weeks after the NYT lawsuit was filed, OpenAI posted a public response to the newspaper in its blogs containing several defenses. For copyright purposes and possible limitations or exceptions, the company pointed out that the memorization results are a programming error in the process of being corrected, that the company applies an opt-out policy that allows proprietary data sources to be excluded from the data sets and, as the main point of controversy, that AI training qualifies as fair use in the U.S. legal system. On this point, OpenAI refers to statements from relevant copyright authorities and stakeholders that have advocated in favor of AI training as fair use. The company concluded its statement on this point and reiterated the correlation between AI and innovation by stating that the, “legal right is less important to us than being good citizens. We have led the AI industry in providing a simple opt-out process for publishers (which The New York Times adopted in August 2023) to prevent our tools from accessing their sites.”

In light of this, it is important to cite the principles on copyright and AI created by the Library Copyright Alliance and the American Library Association:

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* 95

Based on well-established precedent, the ingestion of copyrighted works to create large language models or other AI training databases generally is a fair use. . . . Further, copyright owners can employ technical means such as the Robots Exclusion Protocol to prevent their works from being used to train AIs.

However, these assertions will be tested in the next section.

**APPLYING THE TESTS: WILL EUROPE AND THE U.S.  
REACH THE SAME CONCLUSIONS?**

Under the premise that, during AI training, plausible infringement to the reproduction, public communication, and transformation right occurs, as affirmed by the NYT and the Authors Guild, the required test is analyzing if training constitutes fair use, as affirmed by Open AI. Two tests will be analyzed: the Berne three-step test and the PANE fair use test.

It was initially believed that fair use was incompatible with the Berne test, as it was too broad and uncertain to comply with the special case requirement of Berne. However, international standards have broadened the scope of copyright exceptions in favor of technology and development. For example, an Agreed Statement to the WIPO Copyright Treaty (“WCT”) “reflect[ed] a consensus that existing Berne-compatible L&Es satisfied that test and that suitable new exceptions would as well. It is thus fair to infer that the Agreed Statement serves as a kind of savings clause for L&Es already in existence, including U.S.-style fair use provisions . . . .”

The two separate tests will instead be performed to shed light on the differing rationale behind strict interpretations of the Berne three-step test typical of European systems, based on moral rights, compared to the copyright-based American system. This is because:

[I]t is clear that the prevailing established limitations and exceptions to copyright within Europe do not offer the inherent flexibility or scope of the fair use exception in the US which would enable them to achieve an efficient balance between copyright on the one hand and the freedom of expression and the right to information on the other.

### **BERNE 3-STEP TEST**

The three factors contained in Article 9.2 of the Berne Convention, and subsequently carried over to TRIPS Article 13 and WCT Article 10, establish the following criteria:

Certain special cases: Characterizing AI training as a technological process with determined requirements and execution allows for the consideration of a possible exception for this purpose as an acceptable limitation.

Does not conflict with the normal exploitation of the work: According to copyright and tort law, authors are entitled to claim direct damages and loss of profit damages. Under this perspective and, considering the aforementioned infringements that may occur during AI training, it is arguable that the process conflicts with the normal exploitation of the work. For instance, authors may claim that allowing a copyright exception will cause them to lose licensing opportunities, and thus, loss of profit. Other relevant examples relate to the possible infringement of the transformation right. Several derivative works are available to the public without the authors' endorsement and license to create such adaptations, leaving the author with no appropriate remuneration. Given the limited interpretation of the test, an AI training exception does not comply with this criterion.

Does not unreasonably prejudice the legitimate interests of the right holder: This step relates both to the finality of the exception, as well as to the effect it may have

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* 97

on the authors' rights. Hence, this factor must be analyzed two-fold. Firstly, allowing the exception necessitates the weighing of the right holder's interests and the benefits to society that AI training provides. Secondly, the losses that the exception will inevitably cause to the author must not be unreasonably high, e.g., the author should not be prevented from obtaining a profit from their work. Given the recent trends of European countries attempting to adapt copyright law to the digital age, it is plausible to assume that, even under a restrictive interpretation, the benefits provided by unrestricted AI training are reasonable enough to comply with this criterion, especially if compensation mechanisms are available.

All three of the criteria above must be met for the exception to pass the Berne test. Therefore, the exception fails because AI training directly conflicts with the normal exploitation of the work. This perspective coincides with the demands of the European Writers' Council.

## **FAIR USE TEST**

As previously stated, this essay advances the belief that the fair use test complies with Berne standards. Nevertheless, the different rationale behind this test may induce a different result. The pragmatic and utilitarian U.S. copyright law has been regarded as better suited for addressing the challenges of the digital age. The four fair use criteria ("PANE") includes the following:

Purpose of the use: Given that AI companies operate both for profit or merely for educational purposes, this criterion must be evaluated on a case-by-case basis, wherein educational purposes are considered fair. This does not mean that commercial AI training is to be considered as unfair. Instead, it means the other factor must be weighed by the judge to reach a decision.

Amount and substantiality of the work used: Although the precise amount of the copyrighted works used to train AI are unknown, the text and data mining process requires substantial and large amounts of information provided in the data set. As stated in the Authors Guild class action, publicly available data sets contain entire works of fiction and non-fiction protected by copyright. This means that, at least in relation to the reproduction right, all of the work is being copied, and at least some portion is used for public communication and transformation.

Nature of the work used: Considering the information contained in the dataset, which contains both fictional works, such as novels and scripts, as well as non-fiction, including chronicles and press publications, another case-by-case analysis is necessary. Even if the use of factual works such as the NYT's articles are considered fair, memorization and regurgitation of entire news posts should be considered unfair.

Effect on the market: This particular factor is the key to understanding the different rationale behind each test. In *Berne*, unreasonable prejudice is to be evaluated from the perspective of the author. According to the TRIPS panel in 2000, "prejudice to the legitimate interests of right holders reaches an unreasonable level if an exception causes, or has the potential to cause, an unreasonable loss of income to the right holder." The utilitarian view of the U.S. test is broader in its scope and asks the key question of "Does the use of the copyrighted work in AI training replace the original work?" This essay argues that it does not. As long as the necessary restrictions to the AI are implemented by companies, the program will not be able to prevent future sales or licensing of the copyrighted work. A summary created by AI is unlikely to prevent book sales or impede proper licensing agreements for adaptations or public communication.

## CONCLUSIONS

The restrictive interpretation of the Berne test and the utilitarian application of the PANE criteria lead to an ideological crossroads regarding the future of copyright law. What are we to do with this divergence? How will AI training affect the future of copyright law? These are inquiries that European legislation and American case law will answer soon.

Nonetheless, a compromise must be made to solve the challenges presented to both the American and European systems. On one hand, American judges must deal with the uncertainty that comes with the open-ended fair use defense. How far can it go, and how will judges lay reasonable foundations to address this new technology? On the other hand, European legislation must clarify the scope of the data mining exception contained in directive 2019/790. Is AI training considered a data mining process? How can authors properly opt out of the exception according to Article 4 of the directive?

This essay posits that both copyright systems are properly equipped to resolve these questions satisfactorily. This is because the digital age necessitates that the seemingly irreconcilable differences between the two systems are left behind. This does not mean abandoning the flexibility of the utilitarian U.S. system or disregarding the importance of moral rights that are rooted in the European system. Instead, solutions must simultaneously protect rightsholders and enable uses of AI that further contribute to science and devolvement.

To achieve this, there are three minimum standards. The first is simplifying the opt-out system outlined in Directive 2019/790. Standardizing this process will allow authors to implement reasonable measures to exclude their works from the exception. The second is developing contractual or mechanical licensing systems that facilitate



compensation for the works used in AI training. The third is implementing policy measures regarding compliance and data protection by AI companies. Much remains to be said, but there is hope that the aforementioned copyright systems are properly equipped and prepared to handle this advanced technology.

## REFERENCES

- Muhammad Usman Hadi, et al., *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations and Future Prospects*, <https://perma.cc/9NS3-HMS3>.
- Jason Wei, et al., *Emergent Abilities of Large Language Models*, <https://perma.cc/KJ38-3UYB>.
- Steve Blank, *Artificial Intelligence/Machine Learning Explained*, <https://perma.cc/6WU9-W98F>.
- Augusto Cortez Vásquez, et al., *Procesamiento de Lenguaje Natural*, <https://perma.cc/6Q6N-M869>.
- Jenny Quang, *Does Training AI Violate Copyright Law?*, <https://perma.cc/6R8F-7Z2Q>.
- Alex Reisner, *These 183,000 Books are Fueling the Biggest Fight in Publishing and Tech*, <https://perma.cc/FN3V-P7ZD>.
- Faustine Ngila, *AI Companies are too Cheap to Pay for Legit Books*, <https://perma.cc/P3PV-XX7Z>.
- The Ongoing Battle Against OpenAI and the Tumultuous World of AI Book Datasets*, <https://perma.cc/5S6J-A8CN>.
- Capitol Records, LLC v. ReDigi Inc., 934 F. Supp. 2d 640 (S.D.N.Y. 2013).
- Court of Justice of the European Union C-263/18, <https://perma.cc/5VD8-NJEQ>.
- Cheryl Foong, *Artificial Intelligence and the Right to Communicate to the Public in Copyright: Lessons from the United States*, <https://perma.cc/RY3Q-XSM3>.
- Marina Castells I Marqués, *Cocreación Artística entre Humanos y Sistemas de Inteligencia Artificial: Nuevos Desafíos para el Derecho de Autor: Robótica, Inteligencia Artificial, Tecnología*, <https://perma.cc/99NE-BU2L>.

Daniel J. Gervais, *AI Derivatives: The Application of the Derivative Work Right to Literary and Artistic Productions of AI Machines*, <https://perma.cc/44M6-FKTN>.

Authors Guild v. OpenAI Inc., 1:23-cv-08292-SHS (S.D.N.Y. Dec. 04, 2023).

Winston Cho, *Authors Guild Exploring Blanket License for Artificial Intelligence Companies*, <https://perma.cc/52FY-E7VP>.

The New York Times v. OpenAI Inc., 1:23-cv-11195, (S.D.N.Y. Dec. 27, 2023).

Tobitt Charlotte, *Why New York Times Lawsuit Seeks Destruction of OpenAI and Microsoft LLMs*, <https://perma.cc/D8VQ-L77C>.

Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo.

Gabriella Svensson, *Text and Data Mining in EU Copyright Law*, <https://perma.cc/CPR8-P3K7>.

Elena Vicente Domingo & Teresa Rodríguez Cachón, *Minería de Textos y Datos como (Nuevo) Límite al Derecho de Autor*, <https://perma.cc/RSU9-CFJV>.

Pamela Samuelson & Kathryn Hashimoto, *Is the U.S. Fair Use Doctrine Compatible with Berne and TRIPS Obligations?*, <https://perma.cc/UQ8K-F4TJ>.

Orit Fischman Afori, et al., *Pluralism or Universalism in International Copyright Law Information Law Series (INFO)*, <https://perma.cc/Z8EC-H2R>.

Sunimal Mendis, *Copyright, the Freedom of Expression and the Right to Information: Exploring a Potential Public Interest Exception to Copyright in Europe*, <https://perma.cc/VK2X-AZTJ>.

Eduardo Serrano Gómez. *Responsabilidad Civil, Daños Punitivos y Propiedad Intelectual*, <https://perma.cc/8AQ2-GZ49>.

*European Writers' Council (EWC) Second Statement on the AI Act Proposal*, <https://perma.cc/3CAF-65LR>.

*AI Training Through Copyrighted Works as Infringement:  
Perspectives Under the Berne Three-Step Test and the  
PANE Fair Use Test and Plausible Solutions* **103**

OpenAI. *OpenAI and Journalism*,  
<https://perma.cc/4BK9-M7BQ>.

Library Copyright Alliance *Principles for Copyright  
and Artificial Intelligence*, <https://perma.cc/F8U9-Y2EV>.

Matthew Sag, *The Imaginary Conflict Between Fair  
Use and International Copyright Law*,  
<https://perma.cc/6BC3-NHYU>.

WIPO, *Regional Workshop on Copyright and Related  
Rights in the Information Age*, [https://perma.cc/PM24-  
LMXM](https://perma.cc/PM24-LMXM).