

Training of Artificial Intelligence Models and Copyright: Modern Legal and Ethical Challenges

Lika Sajaia

Doctor of Law, Professor
International University of Georgia GIU
New Vision University
lika.sajaia@giu.edu.ge
<https://doi.org/10.61950/giu.4.2026.12123>

Abstract

The paper explores the legal and ethical challenges of protecting copyright in the process of training artificial intelligence models. The research is based on the view that training AI systems creates a risk of violating authors' economic and moral rights.

The paper analyzes the conceptual difference between human creative activity and AI generation, the stages of AI training, and different legal approaches. Particular attention is paid to the US fair use doctrine, EU text and data mining regulations, Japan's liberal approach, and the copyright protection guarantees of the Berne Convention for the Protection of Literary and Artistic Works.

The research shows that the fragmented legal approach existing in the world cannot ensure equal and effective protection of authors' rights. The paper also evaluates protection mechanisms used during AI training, such as licensing, unilateral statements, website terms, and technical protocols.

Keywords: Generative AI, copyright law, AI training, text and data mining, fair use doctrine.

Introduction

The rapid development of artificial intelligence has presented modern copyright law with numerous new challenges. Generative AI systems function based on large-scale data processing and analysis, while their training is associated with large volumes of textual, visual, and with the use of other types of information. A significant part of this data is protected by copyright law, which is why the question of how AI model training complies with the fundamental principles of copyright protection has become relevant.

The relevance of the issue is further increased by the fact that modern legal systems regulate the AI training process differently. Some jurisdictions prioritize the development of technological innovation, while others prioritize stricter protection of authors' rights. As a result, a fragmented and inconsistent legal environment has emerged, which fails to provide uniform and fair protection in the conditions of a global digital space.

The aim of the paper is to analyze the legal and ethical problems related to copyright in the process of training artificial intelligence models, to demonstrate the conceptual distinction between human creative activity and AI generation, and to evaluate the mechanisms for protecting authors' rights.

The research objectives are: Analysis of the difference between the creative processes of artificial intelligence and humans; Discussion of the technical and legal specifics of AI training; Comparative analysis of the legal regimes of various jurisdictions; Evaluation of the Berne Convention's three-step test in the context of AI training; Determination of the effectiveness of existing legal and technical mechanisms for copyright protection.

The methodological basis of the research consists of dogmatic, comparative-legal, and analytical research methods. The paper uses an analysis of international legal acts, judicial practice, academic literature, and regulations of various countries.

The research hypothesis is that the training of generative artificial intelligence, especially for commercial purposes, does not fit within the traditional framework of copyright limitations and requires a unified, harmonized, and technically enforceable legal regulation that ensures effective protection of authors' rights in the conditions of technological development.

1. Human Creativity and AI Generation: The Necessity of Conceptual Distinction

The term "Artificial Intelligence" is associated with an analogue of human intelligence, although AI systems do not possess the consciousness and cognitive skills that are part of the traditional philosophical and psychological understanding of intelligence serves as the basis. Intellect (Lat. Intellectus - understanding, cognition, mind) is a quality of a person that is related to the ability to think and mental development.

Artificial intelligence is a field of computer science that functions on the principles of algorithmic processing and statistical analysis of data. Accordingly, the "intellectuality" of artificial intelligence can be considered a technological simulation. Assigning the sign of intelligence to a data-driven algorithmic forecasting system is more of a marketing calculation than a reflection of reality.

The issue gains particular relevance when we ask the question - can artificial intelligence create art and literature? Before answering this question, it is appropriate to ask a more fundamental question: can any human create art and literature, or does it still require a special talent? Talent does not imply intellect alone. It also includes creative ability, which is a different dimension from intellect. Accordingly, when the mentioned ability is not universal for humans even in general, its attribution to a well-functioning algorithmic mechanism is theoretically inadmissible.

The creative process is not creation out of nothing. Even the greatest creators draw inspiration from the existing cultural and social environment or from existing works of art. Inspiration, just like creativity, is a purely human quality. The source of inspiration changes form in the creator's consciousness under the influence of their mood, experience, moral criteria, or social/cultural frameworks, and a new, original work is born that expresses the author's individuality.

AI systems do not have the capacity for inspiration, just as they do not have the capacity for creativity. An AI's neural network "reads" texts of hundreds of billions of words in the form of statistical parameters, which then serve as the basis for generating new text.¹

Modern artificial intelligence systems learn from billions of textual units and diverse information sources, as a result of which their "knowledge base" quantitatively significantly exceeds the information accumulated by an individual human during their lifetime. AI's capabilities are based not on subjective experience, but on large-scale data processing, which gives it an unprecedentedly wide range of information coverage, though not the understanding, creative ability, and interpretive depth characteristic of human consciousness. The so-called "knowledge base" of artificial intelligence is achieved through the learning/training of AI models.

The AI learning process, structurally, involves three stages. In the first stage, data collection occurs - extracting text from digital sources. An automated system scans billions of internet pages and processes the textual information placed on them. The second stage is training - a neural network processes these textual data to identify statistical patterns and gradually adjusts tens of billions of parameters during the learning process to better predict and interpret information. In the third stage - inference - the model generates new text.

At all three stages, AI models use works protected by copyright, human intellectual products, and through an algorithmic system, they generate new text/products from them—new but not original, because the text/product is entirely based on what already exists and does not contain even a small element of independent creativity. Therefore, we use different terms: humans create and artificial intelligence generates. From this perspective, the creative process carried out by a human and the generation carried out by artificial intelligence should be conceptually distinguished: a human represents the author who creates a creative and subjective intellectual work, while artificial intelligence acts as a data-driven generative system.

Based on all the above, AI cannot be considered an author in the sense of classical copyright law, although the risk of copyright infringement may arise during its operation, especially when the data used for its training includes works protected by copyright law.

2. Different Legal Regime for AI Training

The legal regime related to the training of artificial intelligence models varies significantly across different jurisdictions, and as a result, we get a different balance between the interests of protecting intellectual property and promoting technological innovation.

In US law, the copyright exception "Fair Use Doctrine" is based on a four-factor test and is a subject of intense debate in the context of AI training. In judicial practice, for example, in *Authors Guild v. Google*² transformative use of text was established because Google Books did not replace the texts but created

¹Paul Goldstein & Bernt Hugenholtz, *International Copyright: Principles, Law, and Practice* (4th edn, Oxford University Press, 2019), ch. 9.

²*Authors Guild, Inc. v. Google, Inc.* 2015. United States Court of Appeals for the Second Circuit, 804 F.3d 202, 216 (2d Cir. 2015).

information about them. This Logically, AI training is often viewed as a transformative process that creates not a copy of the text, but a statistical model.

However, this analogy is not entirely identical. In the case of Google Books, text was indexed and remained partially available to the user, whereas AI models transform text data into parameters and then create new text. Generative artificial intelligence systems do not have a similarly grounded claim of transformativeness for copying protected works. Works are fully copied and mechanical encoding occurs in the AI model, such that it does not provide any search mechanism or other functional purpose, let alone criticism or commentary³. In this case, the user is also no longer informed about the original source. The author suffers both economic damage and their personal non-property rights are also infringed. The decision in *Andy Warhol Foundation v. Goldsmith*⁴ further tightened the assessment of "transformative use," emphasizing that a commercial purpose and a different nature of use are not always sufficient to prove the fair use doctrine.

In the European Union, Directive 2019/790 on Copyright and the Digital Single Market⁵ introduced exceptions for text and data mining. Analysis of text and data by automated methods for scientific research is permissible, while commercial use is permissible only if the rights holder has not used an "opt-out" mechanism (refusal expressed in a machine-readable format). The European Union Regulation on Artificial Intelligence obliges AI companies to disclose general sources of training data, which creates a new transparency mechanism, although it does not provide compensation for past use.⁶

In Great Britain, the Copyright, Designs and Patents Act⁷ allows text and data mining only for scientific research and not for commercial use. In Japan, one of the most liberal regimes operates. Japanese copyright law⁸ allows data analysis and AI training, including for commercial purposes, without the author's permission, if the use occurs within the scope of information analysis.

Accordingly, the legal regime for AI training is globally heterogeneous and regulates the constant balance between the protection of intellectual property and the development of innovation differently.

3. Berne Convention for the Protection of Literary and Artistic Works and AI Model Training

The use of copyrighted works in the AI training process must be evaluated under the Berne Convention for the Protection of Literary and Artistic Works⁹. Because it is the main international act in the field of copyright law. The Berne Convention's three-step test establishes three criteria: the use must be a special

³Jacqueline C. Charlesworth, *Generative AI's Illusory Case for Fair Use*, *Vanderbilt Journal of Entertainment & Technology Law* 27, no. 3 (May 2025): 359.

⁴*Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*. 2023. Supreme Court of the United States, 598 U.S. 508, 532 (2023).

⁵EU Directive 2019/790 (DSM Directive)

⁶Regulation (EU) 2024/1689

⁷UK, Copyright, Designs and Patents Act 1988

⁸Japanese Copyright Act, Article 30-4

⁹Berne Convention for the Protection of Literary and Artistic Works, 1886

case, it must not conflict with the normal economic exploitation of the work, and it must not unreasonably prejudice the legitimate interests of the rights holder.¹⁰

According to this framework, AI training can partially be considered a technologically defined "special case," although the second and third criteria are particularly problematic. During the training of AI models, mass processing of works occurs, which potentially harms their normal commercial use. When AI learns from authors' works en masse and creates new competitive products from them, the author's property rights are violated, and accordingly, they suffer economic damage.

An important theoretical issue in this discussion is the compatibility of the Berne Convention with technological progress. Against the backdrop of technological development and especially generative AI, the question often arises: how adequately does this convention respond to modern digital reality, and does it need revision to better reflect the specifics of data-driven automated systems? According to the Berne Convention, an author's property rights are not absolute and can be restricted. However, such restrictions always serve specific legitimate goals - the free dissemination of information, criticism, and the development of science and education. Beyond these goals, for example, when artificial intelligence training is directed towards a profit-oriented business model, the question arises as to how far the use of the limitation is justified on the same grounds.

If the main purpose of use is to obtain commercial profit and not to protect the public interest, such use no longer fits into the logic of exceptions provided by the Berne Convention. Accordingly, in such cases, the necessity of obtaining the author's permission and appropriate compensation becomes more clearly defined.

Thus, AI training is difficult to fit within the permissible framework of copyright limitation defined by the Berne three-step test, as it often contradicts the normal exploitation of the work and creates a risk of unreasonable harm, while the purpose of use goes beyond traditionally recognized public interests.

4. Legal Mechanisms for Copyright Protection during AI Model Training

Various legal instruments are used for copyright protection during artificial intelligence training. Among them, the following are important: unilateral declaration, license agreement, and website terms and conditions.¹¹

The main problem with legal mechanisms is that in many cases they cannot ensure the perception and automatic execution of the copyright holder's will at a technical level. As a result, they sometimes remain as declaratory and individual instruments rather than systemic protection mechanisms.

A unilateral declaration represents the publicly expressed will of the copyright holder regarding the use of their work. Such a declaration can be placed on a website, attached to a copy of the work, or provided directly to the artificial intelligence provider. However, its effectiveness is significantly limited. A

¹⁰World Intellectual Property Organization (WIPO), Guide to the Copyright and Related Rights Treaties Administered by WIPO and Glossary of Copyright and Related Rights Terms (Geneva: WIPO, November 2003), 56.

¹¹Lucchi, *Generative AI and Copyright*, 2025, 41.

unilateral declaration is primarily human-readable text and does not constitute a computer-understandable instruction. Its enforcement depends on whether a third party perceives it as a legally binding requirement. A unilateral declaration cannot satisfy the criteria of computer readability and automatic executability. Accordingly, it represents an expression of a public position rather than an effective technical or legal barrier.

A licensing agreement is a relatively strong mechanism, as it creates a bilateral, legally binding agreement between the copyright holder and the user. Through it, the terms of use of the work can be defined in detail, including use for AI training purposes and the rules and conditions for issuing appropriate compensation. However, from a practical point of view, the large-scale use of this mechanism is problematic. Training generative AI systems requires huge volumes of data, which implies the necessity of concluding individual or collective agreements with thousands or millions of rights holders.

Accordingly, the licensing system remains an effective tool in individual or specialized cases, but not a universal solution for regulating AI training.

Website terms and conditions, which define the terms of use of the site and the rights and obligations of the user, can also be considered as a legal mechanism for protecting the author's rights. Its effectiveness is also limited. First of all, terms and conditions are mostly written in human language and are not technically structured so that AI systems can automatically perceive their content. This mechanism also fails to create a sufficiently effective and technologically enforceable protection system in the context of AI training.

All three legal mechanisms (unilateral declaration, licensing agreement, and website terms and conditions) play a certain role in the copyright protection system, although their effectiveness is limited due to the specifics of generative AI training. Existing legal mechanisms cannot create a universal, unified, and practically effective framework for protecting copyright in the conditions of AI training. Their main limitation is precisely that they are either too declaratory in nature or are associated with overly complex and costly administrative processes, which is why their effective enforcement is not possible.

It is necessary to develop such a flexible legal model that simultaneously ensures the recognition of rights and their technical enforcement. Only an integrated approach of legal and technical instruments can become the basis for universal protection, where the author's will is not only formally recorded but also actually and automatically implemented by artificial intelligence systems.

Research Results

The development of generative artificial intelligence has presented copyright law with new challenges. Modern AI systems are based on large-scale data processing, and their training is practically impossible without the use of copyrighted works, which is directly related to the protection of the author's economic and moral rights.

The work has shown that human creative activity and generation performed by AI are essentially different processes. A human creates a work based on consciousness, experience, and creative freedom, while AI

functions on the principles of algorithmic and statistical processing. Accordingly, AI cannot be considered an author in the sense of classical copyright law.

The research also confirmed that there is no uniform legal regime for AI training in the world. The USA, the European Union, and Japan use different approaches, which creates a risk of legal uncertainty and unequal protection. It is necessary to create a more harmonized international framework.

Analysis of the three-step test of the Berne Convention showed that AI training cannot be considered a basis for copyright limitation. The limitations allowed by the Berne Convention have historically served the purposes of education, science, and the dissemination of information, while the functioning of generative AI systems is often oriented toward commercial benefit. Accordingly, in such conditions, it becomes particularly important to obtain permission from the author and appropriate compensation in compliance with the law.

Conclusion

For the protection of copyright, it is necessary to create such a flexible and integrated legal model that will combine legal and technical mechanisms and ensure not only the formal recording of the author's will but also its automatic and technical enforcement by AI systems.

The development of artificial intelligence should not occur at the expense of human creative labor and copyright. Technological progress will only be fair and sustainable if it equally protects both the development of innovation and the economic, moral, and ethical interests of authors.

Bibliography

1. Goldstein, Paul, and Bernt Hugenholtz. International Copyright: Principles, Law, and Practice. 4th ed. Oxford: Oxford University Press, 2019.
2. Charlesworth, Jacqueline C. "Generative AI's Illusory Case for Fair Use." *Vanderbilt Journal of Entertainment & Technology Law* 27, no. 3 (May 2025).
3. Lucchi, Nicola. "Generative AI and Copyright." 2025.
4. World Intellectual Property Organization (WIPO). Guide to the Copyright and Related Rights Treaties Administered by WIPO and Glossary of Copyright and Related Rights Terms. Geneva: WIPO, November 2003.

International Acts

1. Berne Convention for the Protection of Literary and Artistic Works, 1886.
2. Regulation (EU) 2024/1689 (Artificial Intelligence Act).
3. Directive (EU) 2019/790 (Digital Single Market Directive).

Laws and Court Decisions

1. Copyright, Designs and Patents Act 1988 (United Kingdom).
2. Japanese Copyright Act.
3. Authors Guild, Inc. v. Google, Inc., 804 F.3d 202, 216 (2d Cir. 2015).
4. Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 532 (U.S. 2023).