

Copyright in Generative Artificial Intelligence

Mário Slomp Filho 

PPCJ—Programa de Pós-Graduação em Ciência Jurídica (Graduate Program in Legal Science), Universidade do Vale do Itajaí, Itajaí, Brazil

Email: marioslomp@gmail.com

How to cite this paper: Slomp Filho, M. (2026). Copyright in Generative Artificial Intelligence. *Beijing Law Review*, 17, 1-10. <https://doi.org/10.4236/blr.2026.171001>

Received: October 24, 2025

Accepted: January 4, 2026

Published: January 7, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Contextualization of the Theme: The rise of generative artificial intelligence (AI) has brought profound changes to modern life, permeating various spheres, from entertainment to science and technology. The ability of these AIs to create original and complex content, in many cases indistinguishable from that produced by humans, raises significant legal questions, particularly regarding copyright. However, current legislation often proves to be inadequate or insufficient to anticipate and regulate all situations emerging from this new technological scenario. **Objective:** This article aims to define the extent to which copyright laws should be applied to content used in the training of generative artificial intelligences, as well as to the content generated by them. The investigation focuses on identifying specific moments when legal protection should be considered, allowing for the protection of creators' rights and stimulating technological innovation. **Methodology:** Regarding the methodology used, in the research phase, the inductive method was employed, in the data processing phase, the Cartesian method was used, and in the results report, the inductive logical basis was applied. **Results:** This study shows that copyright should not apply to content used to train generative artificial intelligence. In a similar way to human beings, who use all the learning accumulated throughout their lives as a reference for their creations without suffering restrictions from copyright legislation, artificial intelligence, developed with the aim of replicating human skills, should also not be limited in relation to the content used for its training. On the other hand, copyright legislation must be applied to content generated by these artificial intelligences. The material produced, which may include text, images, music, and other forms of creative expression, must be protected to ensure that the rights of creators are respected. In this way, this article contributes to the contemporary legal debate, proposing a balanced approach that considers both the need for copyright protection and the importance of encouraging innovation and technological development in the field of artificial intelligence.

Keywords

Artificial Intelligence, Copyright Law

1. Introduction

The development and widespread adoption of generative artificial intelligence (AI) have brought about significant transformations across various spheres of human activity, ranging from the creation of artistic content to the execution of complex scientific tasks. Within this broader technological landscape, Luger (2013: p. 1) defines artificial intelligence as “the branch of computer science concerned with the automation of intelligent behavior”, a characterization that underscores the capacity of AI systems to perform functions traditionally associated with human cognition. This conceptual framework is particularly relevant when examining generative models, whose ability to produce original outputs, frequently indistinguishable from works authored by humans, intensifies the legal debates surrounding authorship and copyright protection. As a result, it becomes increasingly evident that existing legislative frameworks are often inadequate to address the range of unprecedented scenarios arising from these technological advances.

This study aims to investigate whether copyright laws should apply to the content used in the training of generative artificial intelligence systems and, if so, at what stage such legal provisions should be enforced.

The absence of specific legislation to regulate the use of generative AI represents a significant legal and ethical challenge, as this technology is already directly affecting the way content is produced and consumed. The lack of clear norms generates legal uncertainty for both AI developers and content creators, fostering an environment prone to insecurity and potential litigation.

The central hypotheses of this research are as follows: first, that copyright legislation should not apply to the content used to train generative artificial intelligences, given the similarity between this process and human learning, which draws upon accumulated experiences and references without legal restrictions; and second, that the content generated by such technologies should indeed be subject to copyright protection, thereby safeguarding the originality and value of creative works and ensuring the rights of human authors.

In this context, this article is limited to the analysis of the content used to train generative artificial intelligences and the outputs produced by such technologies. Other forms of AI that may involve the use of patented technologies without authorization from patent holders are beyond the scope of this study. This delimitation is necessary to maintain focus on the most pressing and specific legal issues related to generative AI.

From a methodological standpoint, this study adopts a structured and sequential approach. In the initial investigative phase, an inductive method was employed, allowing broader conclusions to emerge from the examination of concrete

cases, normative frameworks, and relevant scholarly contributions. During the analytical phase, the research followed a Cartesian method, decomposing complex legal and technological issues into smaller conceptual units to enable a clearer and more systematic assessment. Finally, the presentation of results reflects an inductive logic, progressively articulating the findings derived from the research and grounding the overarching conclusions in the evidence gathered throughout the study.

Recognizing the need for a critical and well-founded examination of copyright application in the context of generative AI, this study represents a relevant step toward constructing a coherent regulatory framework. By exploring the limits and possibilities of current legislation, this research seeks to contribute to the ongoing legal debate, proposing balanced solutions that protect creators' rights while fostering technological innovation.

2. Human Cognitive Processes

The famous phrase "*In nature, nothing is created, nothing is lost, everything is transformed*" is attributed to the French chemist Antoine Lavoisier and encapsulates the fundamental principle of the conservation of mass in chemistry. Although originally formulated within the context of the natural sciences, this statement carries a broader significance and can be interpreted and applied across various fields of human knowledge.

In the legal sphere, particularly within the domains of intellectual property and copyright law, the phrase may be reinterpreted as a reflection on the notions of originality and the transformation of ideas. Just as in nature nothing is created from nothing, in the creative and intellectual realm, new ideas and works are often built upon preexisting creations.

In this context, the process of knowledge acquisition is influenced both by the individual's cognitive structures and by their interactions with the object of study, as well as by the conflicts that arise from these interactions. When the concepts involved are more complex than those already present within the individual's cognitive frameworks, or when there are no similar references stored in memory, a state of imbalance occurs within these cognitive structures. This imbalance enables the necessary reorganization that leads to the construction of new knowledge (Darsie & Marchiori, 2023: p. 6).

For this reason, everyday problems, intellectual challenges, and the constant imbalance between what is already known and what one seeks to learn are the primary driving forces of learning.

Learning, therefore, unfolds through fundamental stages: disequilibrium, assimilation, accommodation, and equilibrium. When confronted with something unknown, the individual experiences a state of disequilibrium, which may lead either to denial or to curiosity. Curiosity, in turn, motivates the individual to explore and understand the new, thus facilitating learning. Assimilation takes place when new knowledge can be integrated into pre-existing cognitive structures, whereas accommodation requires the modification of prior knowledge in the ab-

sence of relevant references. The continuous interaction between assimilation and accommodation, referred to as “*progressive equilibration*”, enables the evolution of knowledge (Darsie & Marchiori, 2023: p. 7).

The association of ideas in the human brain occurs through biological mechanisms that activate networks of neurons distributed throughout the brain. Memory plays a crucial role in this process, particularly two types: semantic memory and episodic memory. Semantic memory refers to the general knowledge we possess about the world, whereas episodic memory involves specific events that we have personally experienced. These forms of memory are stored and retrieved through processes known as encoding, consolidation, and retrieval, all of which depend on the neurons’ capacity to connect and reorganize among themselves (Eichenbaum, 2017: pp. 19-45).

Beyond biological aspects, psychological factors such as attention, motivation, and emotions significantly influence how we associate ideas. Selective attention enables us to focus on certain stimuli, linking them to relevant memories. Motivation directs us to seek associations that align with our goals. On a more abstract level, the formation of associations between ideas is also shaped by preexisting cognitive schemas, organized mental structures built upon past experiences and accumulated knowledge. These schemas assist in interpreting new information and integrating it into the broader networks of knowledge we already possess (Gazzaniga, 2009: pp. 1047-1048).

In this context, Munari (1998: p. 12), author of “*Design as Art*” (original title: “*Das coisas nascem as coisas*”), criticizes a method widely adopted in design schools that encourages students to constantly seek new ideas, as if it were necessary to reinvent everything on a daily basis. According to him, such an approach does not contribute to the development of professional discipline among young designers; rather, it misguides them and leaves them ill-prepared to face the challenges of the professional market.

Thus, ideas emerge not only from personal experiences but also from the accumulation of knowledge and the observation of other ideas. This continuous process of learning and association is mediated by biological, psychological, and cognitive mechanisms that interact to form a complex network of neural connections. For this reason, every new idea is, in essence, a recombination of pre-existing information, reorganized in a novel way.

3. Artificial Cognitive Processes

Before examining the operational dynamics of Artificial Intelligence in greater depth, it is necessary to recall certain foundational considerations that support a more precise understanding of the topic. The introductory section has already referenced Luger’s classical definition of AI as “the branch of computer science concerned with the automation of intelligent behavior”, a formulation that provides an essential conceptual anchor for the present analysis.

This definition, however, naturally prompts further inquiries: what constitutes

intelligence in the first place, and how should this notion be interpreted when ascribed to computational systems? These questions are not merely semantic; they shape the theoretical background against which contemporary debates on AI and its legal implications must be assessed.

The complexity and vitality of the human mind make it difficult to define “intelligence” precisely enough to evaluate whether a computer program can truly be considered intelligent. The very attempt to define intelligence, whether artificial or innate, raises numerous questions. Is intelligence a single faculty, or a collection of distinct and unrelated capacities? To what extent is intelligence learned rather than innate? What actually occurs during the learning process? What constitutes creativity and intuition? Can intelligence be inferred from observable behavior, or does it require evidence of a specific internal mechanism? These are questions that continue to challenge the field of Artificial Intelligence (Luger, 2013: p. 1).

On the other hand, Blackburn (2016: p. 33) sheds light on the subject by defining “artificial intelligence” in his Dictionary of Philosophy as “the science of making machines that can do the kinds of things that humans can do”. This definition is, quite clearly, both broad and intuitive. It captures the essence of artificial intelligence by emphasizing its interdisciplinary nature, which crosses the boundaries of computer science, psychology, philosophy, and neuroscience.

Thus, Blackburn’s definition underscores the central goal of artificial intelligence: to imitate or replicate the human capacity to think, learn, and solve problems.

In this regard, according to Russell and Norvig (2013: pp. 26-27), authors in the field of computer science, when analyzing a cognitive model approach, one observes the need to understand human mental processes through methods such as introspection, psychological experimentation, and brain imaging, with the aim of developing computational models that simulate the human mind.

To this end, Luger (2013: p. 1) conceives artificial intelligence as a branch of computer science, and therefore grounded in solid theoretical and applied principles. This encompasses the use of data structures to represent knowledge, the application of algorithms to make use of that knowledge, and the employment of programming languages and techniques for its implementation.

In other words, in a more didactic explanation, data in Artificial Intelligence serve essentially the same role as a person’s repertoire: they constitute the totality of information, experiences, and knowledge accumulated over time. Just as a human being learns from what they see, read, and hear, machines learn from datasets that feed their systems.

Conversely, an algorithm in AI functions as a set of instructions or parameters that guide how the machine should process and interact with that data. It is akin to a recipe that defines how information must be analyzed, compared, and applied to perform specific tasks.

Russell and Norvig (2013: pp. 53-54) explain that computer science, over the past six decades, has primarily focused on the study of algorithms. However, it has now become evident that, in many cases, it is more effective to focus on the data

while being less strict about the specific algorithm to be applied. This shift is largely due to the increasing availability of vast amounts of data, such as trillions of English words, billions of web images, or genomic sequences.

For this reason, the authors cite a study on word sense ambiguity, in which they observed that differentiation could be achieved with an accuracy exceeding 96% without the need for labeled examples. Instead, a large body of unannotated text combined with the dictionary definitions of the two senses proved sufficient to label examples and, from there, to train new models capable of labeling additional instances. Indeed, it was found that such techniques improve even further as the amount of available text increases, from one million to one billion words (Russell and Norvig, 2013: pp. 53-54).

In the same line of thought, the aforementioned authors present another study addressing the problem of filling in missing areas of a photograph. Consistent with the results obtained in linguistic tasks, it was found that the performance of the algorithm improved significantly when the image dataset increased from 10,000 to two million photographs.

In other words, these studies suggest that the limitation of available knowledge for AI can be overcome in many applications through learning-based methods, provided that the algorithms designed to expand such knowledge have access to a sufficient amount of data to operate effectively.

Therefore, the functioning of these AI technologies fundamentally depends on the accumulation of data. Consequently, a contemporary debate has emerged regarding the need to regulate copyright protection over such content, particularly with respect to the material used for training AI systems. With the vast amount of data now available, machines are capable of continuously learning and improving their ability to recognize and respond to information. However, the question of how these data are obtained and utilized, and whether there is a need to impose copyright control over them, has become increasingly pressing.

4. Recent Legal Disputes

In December 2023, *The New York Times* filed a lawsuit against OpenAI and its principal investor, Microsoft, alleging copyright infringement. *The New York Times* accused OpenAI of using its content without authorization to train ChatGPT (Associated Press, 2024).

In its petition, the newspaper complains:

Defendants' unlawful use of The Times's work to create artificial intelligence products that compete with it threatens The Times's ability to provide that service. Defendants' generative artificial intelligence ("GenAI") tools rely on large-language models ("LLMs") that were built by copying and using millions of The Times's copyrighted news articles, in-depth investigations, opinion pieces, reviews, how-to guides, and more. While Defendants engaged in widescale copying from many sources, they gave Times content particular emphasis when building their LLMs—revealing a preference that recognizes

the value of those works. Through Microsoft's Bing Chat (recently rebranded as "Copilot") and OpenAI's ChatGPT, Defendants seek to free-ride on The Times's massive investment in its journalism by using it to build substitutive products without permission or payment ([United States District Court for the Southern District of New York, 2023, Case 1:23-cv-11195](#)).

In a subsequent development, eight newspapers, including the *New York Daily News*, also filed lawsuits against OpenAI for the same reason: the unauthorized use of their content for training AI models. This collective action reinforced *The New York Times's* claims, emphasizing the financial harm caused by the dissemination of journalistic content without proper authorization or compensation ([Ferreira, 2024](#)).

The newspapers argued that the practice of data scraping by AI companies undermines the sustainability of quality journalism, as it reduces the incentive for readers to visit the original news websites. The class action emphasized the need to strike a balance between technological development and the protection of the copyright held by content producers.

In the music industry, Universal Music has also faced challenges of this nature. In a notable case, an AI-generated song mimicked the style of the artist Drake and was distributed across multiple streaming platforms without authorization from the record label or the artist himself. Universal Music argued that the reproduction and distribution of musical works without permission constitute a clear violation of copyright and negatively affect the music industry, which depends on exclusivity and control over the dissemination of its works ([O'Dell, 2024](#)).

This case raised questions about the responsibility of streaming platforms and AI developers in overseeing and controlling the content generated by their technologies. Universal Music demanded that the platforms remove the infringing song and implement stricter measures to prevent the recurrence of such incidents.

In the entertainment field, actress Scarlett Johansson filed a lawsuit against OpenAI, alleging that her voice was used without permission to develop the voice of an AI named *Sky*. Johansson argued that the use of her voice without consent not only violates her rights to image and privacy but could also create confusion and mislead fans and the general public ([Olivieri, 2024](#)).

These recent disputes reveal that the rapid expansion of artificial intelligence technologies has not only redefined the boundaries of creativity and authorship but has also challenged long-established legal and ethical frameworks. It must be noted, however, that many of these cases remain pending or subject to appeal, particularly within the United States, where the judiciary is still in the early stages of delineating a coherent path forward. Thus, the current landscape reflects emerging tendencies rather than a consolidated body of jurisprudence. As society grapples with the implications of using protected content to train generative models, a deeper philosophical question emerges: can machines truly *create* in the same way humans do? To explore this issue, it becomes essential to examine how ideas are generated, both in the human mind and within artificial systems, and to what extent the processes underlying human creativity can be replicated by machines.

5. Idea Generation: Humans vs. Machines

As previously discussed, human beings draw upon all the experiences and knowledge accumulated throughout their lives to create new ideas. In a similar manner, artificial intelligences, designed to emulate this human capability, are trained on vast volumes of data to learn and generate new content. The underlying logic is that if human references are not regulated by copyright law, the same principle should apply to the material used to train AI technologies. From this perspective, the restrictions imposed on AI training data appear inconsistent, since artificial intelligences are merely replicating a natural human process of learning.

A frequent objection to this analogy rests on the assertion that the learning processes of humans and artificial intelligences differ not only in scale but also in nature, particularly because AI systems process information at speeds and levels of systematic organization that far exceed human cognitive capacities. However, this distinction, although empirically accurate, does not inherently negate the structural similarity underlying both forms of learning. The acceleration and automation of informational processing do not, by themselves, transform the act of referencing prior material into a legally distinct category; rather, they merely reflect technological efficiency applied to a function analogous to human cognition. Moreover, it is important to note that, at the current stage of development, AI-generated content emerges only through human interaction, specifically through the formulation of prompts that guide the system's responses. In many cases, the use of protected content during training intersects with the user's own familiarity with such materials, revealing a direct correlation between the system's output and the human repertoire that informs and shapes the eliciting prompt. This dynamic further reinforces the view that the normative evaluation should focus less on the quantitative disparity in data processing and more on the qualitative nature of the output and the presence, or absence, of unlawful appropriation within the generated result. Such an approach preserves conceptual coherence while acknowledging the technological specificities of AI systems without overstating their juridical relevance.

On the other hand, this position is reinforced by a simple observation of the legal proceedings discussed above: in all the aforementioned disputes, it is clear that although the plaintiffs' complaints were directed at the unauthorized use of their content for AI training, the actual infringement occurred in the final output generated by the AIs, that is, texts, voices, and other results identical to those of their original references.

In this sense, by analogy to what applies to human beings, the most coherent approach would be to apply the provisions governing copyright protection solely to the content generated by these new technologies.

6. Final Considerations

The present research sought to examine the application of copyright law to the

material used in training generative artificial intelligences, as well as to the content resulting from the use of such technologies. The relevance of this topic becomes evident in light of the growing use of AI across various spheres of intellectual and cultural production, which raises ethical and legal questions concerning the protection of copyright.

The findings confirmed the initial hypotheses: existing copyright legislation should not apply to the content used for training generative artificial intelligences. This understanding is grounded in consistency with how legal provisions are applied to human creators. Just as humans draw upon their accumulated repertoire of experiences and knowledge as reference points for generating new ideas, AIs rely on data for their training. By the same logic, this principle should be extended to artificial intelligences.

However, it has become clear that the content resulting from the use of these AIs should indeed be regulated under copyright law, in order to protect original creators and ensure fair compensation for their intellectual efforts. In this regard, an important contribution to the contemporary debate is offered by Lemley and Casey (2020: pp. 153-156), who propose a reinterpretation of the fair use doctrine in the context of machine learning through what they call the principle of *fair learning*. According to the authors, when the purpose of the use is not the appropriation of protected elements but rather the extraction of unprotected components, such as facts, ideas, or general structural patterns, the training activity should be deemed presumptively legitimate under the first fair use factor concerning the purpose and character of the use. Even where other factors, such as the nature of the copyrighted work or the amount taken, might suggest an unfavorable outcome, the inherently transformative and learning-oriented character of AI training would justify recognizing the lawfulness of such uses. This perspective further reinforces the distinction between training inputs, which should not be subject to copyright restrictions, and the outputs generated by AI systems, which must be evaluated under the traditional standards of copyright law.

The study demonstrated that the current regulatory framework is not yet adequately equipped to address the particularities inherent in the field of generative artificial intelligence. Therefore, further research on these issues is recommended, with a particular focus on civil liability in cases of copyright infringement caused by the use of AI technologies.

Finally, it is essential that legislators consider the creation of a legal framework adapted to digital realities and to the new forms of creation and distribution of content enabled by AI systems. Such a framework must be sufficiently flexible to keep pace with the rapid evolution of technology without losing sight of the fundamental principles that underpin the protection of copyright.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Associated Press (2024). *Microsoft, OpenAI Hit with Copyright Infringement Lawsuit from the New York Times*.
<https://www.wfaa.com/article/news/nation-world/new-york-times-sues-open-ai-microsoft/507-92912f68-e396-44d6-a561-ca095e364133/>
- Blackburn, S. (2016). *The Oxford Dictionary of Philosophy* (3rd ed.). Oxford University Press. <https://doi.org/10.1093/acref/9780198735304.001.0001>
- Darsie, M. M. P., & Marchiori, S. T. (2023). A concepção construtivista da aprendizagem como contribuição no processo de avaliação da educação básica da aprendizagem como contribuição no processo de avaliação da educação básica. *Contribuciones a las Ciencias Sociales*, 16, 26243-26258. <https://doi.org/10.55905/revconv.16n.11-088>
- Eichenbaum, H. (2017). Memory: Organization and Control. *Annual Review of Psychology*, 68, 19-45. <https://doi.org/10.1146/annurev-psych-010416-044131>
- Ferreira, T. (2024). *OpenAI recebe mais um processo por violação de direitos autorais*. <https://olhardigital.com.br/2024/05/01/pro/openai-recebe-mais-um-processo-por-violacao-de-direitos-autorais/>
- Gazzaniga, M. S. (2009). *The Cognitive Neurosciences* (4th ed.). MIT Press.
- Lemley, M. A., & Casey, B. (2020). Fair Learning. *SSRN Electronic Journal*, 101-181. <https://doi.org/10.2139/ssrn.3528447>
- Luger, G. F. (2013). *Artificial Intelligence* (6th ed., D. Vieira, Trans.). Pearson Education do Brasil.
- Munari, B. (1998). *Das coisas nascem as coisas* (J. M. de Vasconcelos, Trans.). Martins Fontes.
- O'Dell, B. (2024). *Spotlight on Large Language Models: Copyright Law in the Age of Digital Creativity*. <https://vlex.com/blog/writing-competition-2024-llms-copyright?utm/>
- Olivieri, F. (2024). *Scarlett Johansson acusa OpenAI de usar voz semelhante à dela no ChatGPT*. Exame. <https://exame.com/inteligencia-artificial/scarlett-johansson-acusa-openai-de-usar-voz-semelhante-a-dela-no-chatgpt/>
- Russell, S., & Norvig, P. (2013). *Artificial Intelligence: A Modern Approach* (3rd ed., R. C. Simille, Trans.). Elsevier.
- United States District Court for the Southern District of New York (2023). *The New York Times Company v. Microsoft Corporation, OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation LLC, and OpenAI Holdings LLC, Case No. 1:23-cv-11195*. <https://storage.courtlistener.com/recap/gov.uscourts.nysd.612697/gov.uscourts.nysd.612697.1.0.pdf>