



GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY: G
INTERDISCIPLINARY

Volume 24 Issue 2 Version 1.0 Year 2024

Type: Double Blind Peer Reviewed International Research Journal

Publisher: Global Journals

Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Artificial Intelligence-Generated Art works: Protection of Copyright in Brazil for Artistic Works used in AI Databases in Light of the U.S. Fair use Doctrine

By Amanda Veríssimo Almeida Vale & Éfren Paulo Porfírio de Sá Lima

Federal University of Piauí

Abstract- This study analyzes the applicability of the fair use doctrine in the context of using artistic works in databases for training Generative Artificial Intelligences (AI). It employs bibliographic research combined with a qualitative-quantitative approach to cases filed in U.S. courts concerning copyright issues involving generative AIs. Initially, the discussion centers on the protection of intangible assets and the intensification of the dematerialization of property in the informational society. The study then focuses on identifying the guidelines provided by copyright theories and their compatibility with works generated by Artificial Intelligence (AI).

Keywords: *copyright. digital goods. artificial intelligence. database. machine learning. utilitarianism. fair use. changes in private order.*

GJCST-G Classification: *ACM Code: K.5.2, H.3.4, I.2.4*



ARTIFICIALINTELLIGENCEGENERATEDARTWORKSPROTECTIONOFCOPYRIGHTINBRAZILFORARTISTICWORKSUSEDINAI DATABASESINLIGHTOFTHEUSFAIRUSEDOCTRINE

Strictly as per the compliance and regulations of:



RESEARCH | DIVERSITY | ETHICS

© 2024. Amanda Veríssimo Almeida Vale & Éfren Paulo Porfírio de Sá Lima. This research/review article is distributed under the terms of the Attribution-NonCommercial-No Derivatives 4.0 International (CC BYNCND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Artificial Intelligence-Generated Art Works: Protection of Copyright in Brazil for Artistic Works used in AI Databases in Light of the U.S. Fair use Doctrine

Amanda Veríssimo Almeida Vale ^α & Éfren Paulo Porfírio de Sá Lima ^ο

Abstract- This study analyzes the applicability of the fair use doctrine in the context of using artistic works in databases for training Generative Artificial Intelligences (AI). It employs bibliographic research combined with a qualitative-quantitative approach to cases filed in U.S. courts concerning copyright issues involving generative AIs. Initially, the discussion centers on the protection of intangible assets and the intensification of the dematerialization of property in the informational society. The study then focuses on identifying the guidelines provided by copyright theories and their compatibility with works generated by Artificial Intelligence (AI). Finally, it explores the contrast between the use of protected works by AI programs and the American doctrine of fair use, highlighting the utilitarian perspective in this debate and the applicability of this doctrine in Brazilian law. Based on the proposed analysis, the investigation concludes that adopting an extensive interpretation of Brazilian copyright law is more advantageous to facilitate the influx of new contributions from the real world. Consequently, it becomes necessary to analyze the applicability of fair use by Brazilian courts and weigh the arguments for and against considering the creation of databases with protected works as fair use.

Keywords: *copyright. digital goods. artificial intelligence. database. machine learning. utilitarianism. fair use. changes in private order.*

I. INTRODUCTION

The relevance of the discussion about generative AIs is especially evident due to the exponential growth in the amount of data feeding these systems. The report on advances in Artificial Intelligence prepared by Stanford (Brynjolfsson *et al.*, 2023, p. 54) revealed that over time, the development of data interpretation parameters showed a regular advance that was particularly pronounced in the 2010s. The complexity of the criteria provided reflects the breadth of tasks that the machine can perform.

In the field of discussions involving copyright protection, as well as in labor rights, the massification of generative AIs had an impact similar to that observed in

Author α: Bachelor of Law from the Federal University of Piauí. e-mail: amandaverissimov@gmail.com

Author ο: PhD (*summa cum laude*) in Private Law from the University of Salamanca, Spain (2016). Degree from Salamanca (*summa cum laude*) in Private Law (2011). Advanced Studies Diploma from the University of Salamanca, Spain (2011). Associate Professor of Civil Law and Permanent Professor of the Graduate Program in Law at the Federal University of Piauí, Brazil. e-mail: efrencordao@ufpi.edu.br. ORCID: <https://orcid.org/0000-0002-0731-4796>.

the early 19th century. On one hand, proponents of the new discovery argue that copyright would be an obstacle to technological advancement and overcoming humanism; on the other hand, the more measured advocates support protecting artists' copyrights to prevent the illicit enrichment of users and the companies owning AI algorithms.

The first position is cautious regarding the protection of the interests of those involved in creative processes. However, this way of framing the issue leads to reasoning about the extent to which individual wills can or cannot override the collective interest. The legal dilemma thus unfolds into a moral problem of a utilitarian nature (Tenório Filho & Mallmann, 2018).

The meaning and scope of the applicability of fair use, in this sense, must be considered from the comparison of favorable and unfavorable perspectives, taking into account the public interest and economic development in conflict. Similarly, the guiding question of this study is presented: what is the applicability of the fair use doctrine in the case of copyrighted works in generative AI databases?¹

To answer this question, the development of the work will follow three stages. Firstly, to analyze the dematerialization of property in the face of new technologies proposing the virtualization of this doctrine, it will be necessary to outline a brief historical overview of the emergence of intangible asset protection and discuss the greater protection granted to them in the context of the information society based on the lessons taught by Fernando Taveira Jr. (2015, p. 49).

In the second stage, it proposes identifying the guidelines provided by copyright theories by analyzing what guarantees should reach works made by Artificial Intelligence, highlighting the questions raised by Samuelson (2023).

Finally, it seeks to discuss the arguments listed by the doctrine to fit the use of artistic works serving the AI program's database into the fair use criteria. For this purpose, the study draws on the reflections of Tenório

¹ This study is the result of a monograph originally presented in Brazilian Portuguese by Amanda Veríssimo Almeida Vale, under the guidance of Éfren Paulo Porfírio de Sá Lima, during the Bachelor of Law course at the Federal University of Piauí, Brazil, having been approved with the highest grade by the evaluation committee of the Final Course Work.

Filho and Mallman (2018), in addition to scrutinizing the understandings of the U.S. courts for the application of the doctrine.

To achieve the proposed objectives of this study, which initially employs a primarily qualitative approach, the methodology used was conceptual analysis through bibliographic review, following an analytical path of academic texts, books, legal documents, and jurisprudence. As pointed out by Gil (2017), the stages of the review include preliminary bibliographic survey, source search, note-taking, and final text writing. This methodology is justified given the peculiarities of the guiding problem chosen as the research object.

However, the course of the work is carried out from a qualitative-quantitative approach to the study object. While purely quantitative methodology prioritizes data collection and systematization, "in qualitative research, especially those where a theoretical analysis model is not previously available, it is common to see a back-and-forth between observation, reflection, and interpretation as the analysis progresses" (Gil, 2017, p. 90). In the first stage, this research qualitatively develops the proposed categories to, in the final stage, make a quantitative survey of cases filed in American courts on copyright matters involving generative AIs.

Based on the aforementioned theoretical framework, the article aims to scrutinize the conflict between the phenomenon of creating artistic products by AI and the applicability of copyright law. By the end of the research course, it is expected to analyze the relationship between the AI product and the work that feeds the database, in light of the fair use doctrine criteria.

II. THE DEMATERIALIZATION OF PROPERTY

In the context of the Information Society, intangible assets gain significant relevance as property gradually loses its material form due to the digitalization of assets (Taveira Júnior, 2015, p. 49). This context involves a paradigm shift where 'liquidity,' as proposed by Zygmunt Bauman (2001, p. 173), points to a scenario in which immaterial objects become the main sources of profit on an ever-expanding scale.

Digital goods can be defined as information stored on electronic devices and the internet. This definition is intentionally imprecise, as it should not only serve as a benchmark for defining the contours of this phenomenon currently but also comprehend present and future innovations. In the current context, it is noted that the phenomenon of the dematerialization of property is closely related to digitalization. This alteration from analog to digital can result from both the transformation of physical documents into electronic files and the direct creation of documents in digital format.

Moreover, the massive circulation of digital goods has been favored mainly by the possibility of greater production and dissemination of content by users and the growing data storage in so-called cloud computing (Taveira Júnior, 2015, p. 55). These circumstances have also made the relationship between technology companies and users of these services more complex.

The deepening of this relationship reveals the user's vulnerability concerning the protection of intangible assets situated in the digital realm. Thus, the internet becomes considered a new field over which Civil Law must focus, especially to correct asymmetry and balance relationships in the digital space. Hence, there arises the need to establish mechanisms to better regulate and manage digital goods.

Indeed, digital goods should also be considered as objects of legal protection. Although part of the doctrine argues that the protection of these goods does not fit within the classic notion of property, it is indisputable that goods of this nature have both property and extra-property value, and thus require regulatory parameters. Currently, digital goods can represent the primary source of income for some users, as profiles and publications on social networks have added value and can be monetarily assessed².

In the digital context, the faculties of use and enjoyment prevail, as access coupled with the possibility of sharing proves more valuable than simply holding ownership. For this reason, there is a greater concern with regulating usage licenses rather than ownership itself, demonstrating the importance of "access protection" (Ehrhardt Jr. & Guilhermino, 2015, p. 2).

It is undeniable that the development of the internet has dynamized the process of data transmission, which can take various formats, such as texts, audio, videos, images, or even works protected by copyright. In digital format, it becomes easier for the author to widely disseminate their creations without major obstacles. Conversely, greater exposure also facilitates the unauthorized use of these works, as the internet provides greater ease of access and dissemination of intellectual creations and the emergence of new uses not covered or foreseen by copyright law (Tenório Filho & Mallmann, 2018, p. 68).

From this dynamism arise important implications in the context of creating digital goods through generative artificial intelligence. Since the operation of this type of AI depends on a reference database for "machine training," it is understood that simple access to these materials and data is already sufficient to meet the purpose of feeding the database.

² According to Marcos Ehrhardt Jr. and Everilda Guilhermino (2015, p. 2): "an application or an account on a social network can represent a much greater monetary value than a tangible asset."

III. ARTIFICIAL INTELLIGENCE AND COPYRIGHT PROTECTION

Artificial intelligence is understood as a system developed to perform specific tasks using coordinated algorithms designed to emulate human capabilities. According to the definition provided by the American National Artificial Intelligence Initiative Act of 2020:

The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to (a) perceive real and virtual environments; (b) abstract such perceptions into models through analysis in an automated manner; and (c) use model inference to formulate options for information or action (GPO, 2020).

In turn, the European Parliament defines AI as the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity (European Parliament News, 2020).

In Brazilian legislation, there is still no specific definition or regulation for this area. However, a commission of jurists was recently established to update the Civil Code in light of the digital revolution, with special concern regarding AI implications (*Senado Notícias*, 2023). There is also a bill in the Senate attempting to establish a regulatory framework for the topic, Project Law No. 5,691/2019 (Brazil, 2019), which outlines guidelines for these systems.

Specifically, generative artificial intelligence differs from other approaches as it is based on a predictive structure of machine "learning." The operation of such systems relies on algorithms of varying complexity levels representing pre-established action chains based on collected data. The quantity and quality of data are significant variables in the resulting product.

Indeed, the AI product may vary drastically depending on the information selected to feed the system. This is because the machine learning technique essentially recognizes patterns in data sets and performs specific tasks "based on its own experience" (Cerri & Carvalho, 2017, p. 298), which is founded precisely on processed data. Using this technique obviates the need for specific programming for each command, as the inserted data are described and the results of their interactions are predicted (Hurwitz & Kirsch, 2018, p. 4). Choosing the data that feeds the system is an essential step to ensure that the results are meaningful and accurate (*idem*, p. 9).

Generative AI algorithms function similarly to human neural networks and are powered by the so-called "Generative Adversarial Networks" (GANs). This technique involves using two types of algorithms: one of a discriminative nature and the other of a generative nature (Goodfellow, 2014, p. 1).

It is evident that applying generative tools allows users to make compositions, including artistic ones, very similar to those of human authorship. For this purpose, these systems are supplied with reference models for the algorithm's "training," necessitating the creation of databases to facilitate data collection and processing.

Thus, the goal is to achieve results close to those stored in the database produced by humans. Given the need to provide a database to enable the operation of AI programs like Stable Diffusion³, the problem lies in the absence of authorization from the authors of the works whose data are used, who are not even recognized with the right of authorship⁴ over the creation nor receive any compensation for the economic benefit derived from the generated work.

Since the reference database with image references is an essential tool for providing new content by the machine, the program may use materials protected by specific licenses, impacting the discussion about the copyright ownership of the AI product.

In this perspective, two systems are highlighted in the context of copyright protection. Typical of Anglo-American law, the "copyright" system of author rights emphasizes the protection of the materiality of the work, with an emphasis on the exclusive right to copy, following the principle "what is worth copying is worth protecting." The French system of "Droit d'Auteur," which has the corollary of protecting the author's figure, influenced the current Brazilian legislation, as evidenced by the personalistic character adopted by Law No. 9,610/98 (Copyright Law or LDA) (Brazil, 1998).

At the international level, the Berne Convention stands out as "the first international and multilateral legal instrument for copyright protection" (Souza, 2006, p. 182), accommodating the two original strands: "copyright" and "Droit d'Auteur," which coexist internationally nowadays.

It is observed that Copyright Law has a hybrid, dual, or "*sui generis*" legal nature. The author is recognized with two bundles of rights. One concerns moral or personal rights (strictly personal) enshrined in Article 24 of the LDA. These act as bonds that unite the author or creator to the intellectual work, protecting the rights inherent to their personality. They arise with the creation of the work, produce perennial effects, and aim to maintain the bond with the created work. On the other hand, there are property rights detailed in Article 29 of

³ About this program, the Stanford Report points out: "Stable Diffusion is trained on existing images created by humans and gives no credit or acknowledgment, leaving open questions around the ethical use of image generators." (Maslej *et al.*, *op. cit.*, p. 77).

⁴ "The property right manifests in two dimensions: a positive and a negative. The first one refers to the claim to mention the author's name. The second one establishes the right to react to violations committed." (Ascensão, *opcit.*, p.141).

the mentioned law, relating to the economic exploitation of legally protected works.

The "Droit d'Auteur" doctrine in its genesis in French law sought to protect creations of the spirit. Thus, the intention was to provide legal support to the author's inspiration as well as the novelty and originality of the created work. In this sense, creativity is considered the main raw material of Copyright Law. Similarly, Leandro Macedo Poli (2008, p. 112) argues that "it is precisely in the creative character of the work that the justification for copyright protection lies."

Indisputably, with the disruptive model proposed by AI, it is necessary to align the Copyright Law theory with new modalities of intellectual work production without, however, interrupting technological development. In the context of generative technologies, the discussion takes on specific contours, and the controversy revolves around the potential violation of copyright due to the program's ability to provide new works similar to those contained in the database used.

In this regard, it is noteworthy that Article 7, item XIII of the LDA provides for the protection of databases provided that "by their selection, organization, or disposition of their content constitute an intellectual creation." However, paragraph 2 of the aforementioned article clearly states that "this protection does not prejudice any copyrights subsisting with respect to the data or materials contained therein." Thus, the copyrights of materials contained in compilations remain in force independently of the database protection itself.

When using pre-existing references in programs applying generative technology, the works generated by the algorithm may reference some protected creation. Therefore, it is necessary to analyze what link is established between the generated work and the one contained in the database from the perspective of investigating a possible violation of the rights of the artists whose works comprise the reference database used by the AI.

In an attempt to answer the question about the possible violation of copyrights by the machine, Beatriz Bugallo Montañó (2022) argues that if the influence of other people's works on the human author's creation is not considered misuse, then composing the database for training the algorithm should be understood similarly, that is, merely as a "source of inspiration." According to the author, the method used by the machine also differs from derivative works. Indeed, subparagraph "g" of item VIII of Article 5 of the LDA states that derivative works result from the transformation of an original work. In this sense, they correspond to the reproduction, modification, and incorporation of a work to generate a new creation distinct from the original and with its own individual content.

Given the presented innovations, doctrinal discussions arise about AI's creative potential, and there seems to be resistance to granting such rights to the

machine. Commenting on the topic, the U.S. registration authority defined that only those works using AI as a mere tool, necessarily requiring human author control, are entitled to copyright protection (United States of America, 2017, p. 68-69).

In another perspective, the problem seems more complicated when considering an unprecedented AI product developed by an AI system with a higher degree of autonomy. In other words, different analysis should be considered when the operator merely provides generic parameters for the production of results over which there is no human control, as in these cases, the machine's role is not merely to obey commands and instructions given by the user.

Restricting the discussion to the context of generative artificial intelligence, copyright protection must also be discussed from the angle of the works used for algorithm training. For some artists, the use of art without consent is an absolute violation of both property and moral rights over the work.

An action is pending in the U.S. courts filed by artists whose works were improperly used as a training base in an AI program (Stable Diffusion Litigation, 2023). The plaintiffs argue that the use of their creations for this purpose was not consented to, exacerbated by the lack of means to allow artists to opt-out of the database.

As noted, AI, although it develops unprecedented products, will provide results similar to pre-existing works, which are somehow ready to compete in the market with the arts used as training data for the algorithm⁵.

In contrast to the artists' plea, it is argued that the use of works is solely to develop and train the algorithm, thus fitting within the so-called fair use. This doctrine, by defining parameters that make the authorization of authors for the use of the work unnecessary, seeks to expand the range of hypotheses for the use of protected material without causing a violation of the creators' rights.

IV. THE FAIR USE DOCTRINE APPLIED TO THE USE OF ARTISTIC WORKS BY ARTIFICIAL INTELLIGENCE PROGRAMS

According to Jeremy Bentham (1974), the most relevant objective is to maximize happiness by safeguarding individuals' pleasure over pain. Utility, in general terms, would be characterized as something that brings happiness and avoids suffering, this being the maxim to be sought by citizens and legislators (Tenório Filho & Mallmann, *op. cit.*, p. 62). Later, to reconcile individual rights with utilitarian ethics, John Stuart Mill (2000) proposes a theory of fundamental

⁵ "Algorithms can be used to create forgeries of existing artworks or to create works that resemble a specific style or artist, which can lead to a devaluation of the art market and a loss of consumer confidence." (Varella, 2023).

rights that mitigates the effects of utilitarian calculation, allowing the articulation between a democratic regime and minimal State intervention, ensuring individual freedom.

The utilitarian thought rejects the notion of natural right, as the goal is to reduce the notions of justice and moral action to calculable precepts, which is why it is incompatible with pre-established aspects. Thus, natural law is replaced by the mentioned idea of utility as the justifying element of law-making. Applying these concepts to the aspect analyzed in the present study, Alves and Pontes (2009, p. 9872) explain:

According to the utilitarian theory, the authors' right corresponds to a prerogative granted to creators so that they can commercially exploit their creation exclusively for a temporary period, justified as an incentive instrument for intellectual creation and authors' remuneration. The natural law theory, on the other hand, is based on the argument that the author has a property right over their work, which is inherent to the creation itself, because it would ensure ownership over the fruit of their labor.

According to the authors, John Locke opposes the utilitarian trend and argues that the author's right should be perpetual, as it holds a moral and personal dimension (Alves & Pontes, 2009, p. 9879). Thus, the work deserves exclusive protection for representing the intellectual labor of its creator. The exclusivity prerogative is seen as an instrument to encourage intellectual production and consequently the authors' remuneration. However, when analyzing copyright legislation under the utilitarian argument, the public utility of disseminating intellectual works to the community is considered, as opposed to the real and commercial monopoly of these works.

In an attempt to align the individual interests involved in copyright protection with the public utility of disseminating protected content, the fair use doctrine, typical of American law, emerges, aiming to limit copyright protection according to some criteria discussed below.

The fair use theory resulted from jurisprudential consolidation in Section 107 of the "Copyright Act" (United States of America, 1976) after a series of decisions intended to establish broad access to works as a way to achieve cultural development and technical-scientific progress without, however, generating risks to the interests of copyright holders.

The purpose of fair use in preventing copyright protection from making all forms of a work's use unfeasible aligns with the utilitarian discussion related to public interest and cultural, economic, and technological development. Thus, fair use defines some criteria for weighing these interests, among which are: (i) the purpose of the use; (ii) the nature of the protected intellectual work; (iii) the amount and substantiality of the portion used; and (iv) the level of harm to the work's author in the market.

These parameters were created in the paradigmatic case "Folsom vs. Marsh" (C.C.D. Mass 1841), in which the applicability of the fair use doctrine was discussed. The dispute involved the copyright of George Washington's writings belonging to the editors who filed the action. The controversy arose from the copying of several excerpts from this text in the creation of a biographical work of the president. In the end, the court found that the defendants' conduct constituted a violation of copyright, noting that use cannot be tolerated when essential parts of the work are used to substitute the original work, as this would discourage future creators. Based on these criteria, it is seen that the fair use theory better adheres to specific uses that include criticism, commentaries, news reporting, research, and classroom teaching (United States of America, 1976, p. 20).

In this perspective, a functionalization of legal institutions – in this specific case, property – is observed in line with the contributions of the Italian philosopher Norberto Bobbio (2007). According to the author, the law also incorporates a promotional function, providing means of inducing and incentivizing certain behaviors, in addition to the roles traditionally attributed to the structure within the legal order. However, Brazilian legislation does not prioritize the incentive for the author's creativity but rather privileges the property aspect related to the work.

In this line, Brazilian copyright law demonstrates a strong protective character and gives little relevance to the weighing of the exclusivity right in face of collective interests, despite the Federal Constitution (CF), in its Article 5, item XXIII, being explicit in determining the observance of the social function of property. Article 46 of the LDA⁶ outlines a list of limitations to copyright. The

⁶ Article 46. The following do not constitute copyright infringement: I - the reproduction: a) in the daily or periodic press of news or informative articles published in daily or periodical newspapers, mentioning the author's name if signed and the publication from which they were transcribed; b) in daily or periodical newspapers of speeches delivered in public meetings of any nature; c) of portraits or other forms of representation of the image made on commission when carried out by the owner of the ordered object, without opposition from the person represented or their heirs; d) of literary, artistic, or scientific works for the exclusive use of the visually impaired, provided that reproduction is non-commercial and made using the Braille system or other procedure in any medium for these recipients; II - the reproduction of small excerpts for private use of the copier, provided it is done by them without intent to profit; III - the citation in books, newspapers, magazines, or any other means of communication of passages from any work for study, criticism, or controversy, to the justified extent for the intended purpose, indicating the author's name and the source; IV - the collection of lessons in educational establishments by those to whom they are directed, provided that publication is prohibited without prior and express authorization from the lecturer; V - the use of literary, artistic, or scientific works, phonograms, and radio and television broadcasts in commercial establishments exclusively for demonstration to customers, provided that these establishments sell the media or equipment that allow their use; VI - the theatrical representation and musical performance when carried out in the family environment or exclusively for educational

concessions made by the Brazilian legislator are insufficient as they cannot cover all situations.

Nonetheless, it is observed from the reading of the items of the mentioned article that the legislator's central concern was to allow use as long as it falls under non-commercial use hypotheses – except for items III and VIII – and is permissive regarding educational and informative use.

According to Schirru (2020, p. 157), the interpretation applied to the list of limitations should be considered extensive, as it enables the incidence of the fair use doctrine under national legislation.

In Brazil, the Superior Court of Justice (STJ) permits the use of some protected works without authorization, provided that the "Three-Step Rule" requirements (Brazil, 2011) established in the Berne Convention are met. This theoretical orientation relates to the interpretation of copyright limitations and outlines criteria for the flexibility of copyright protection: (i) the use must fit special cases; (ii) the use must not conflict with the normal exploitation of the original work; and (iii) no unjustified harm must be caused to the legitimate interests of the right holder. However, it is notable that the criteria established by this theory are quite generic and imprecise, which demonstrates that the fair use doctrine offers a better approach to this issue.

Finally, shifting this discussion to the scope of this investigation, a survey is conducted on the justifications and arguments both for and against the application of the fair use doctrine to the use of protected artistic works to supply the image databases of generative AI programs.

a) *Favorable Arguments for Applying Fair Use*

In the U.S., providing new images through applications has been the subject of legal actions based on allegations of plagiarism and art theft from the images that comprise the database. According to the litigants, the "Stable Diffusion" program, launched by Stability AI, "contains unauthorized copies of millions – and possibly billions – of copyrighted images." In their plea, these artists claim that these copies were made without the authors' knowledge or consent.

According to Pamela Samuelson (2023), an argument used by developers, in contrast to the artists' allegations, is that the technology's use does not focus on the artistic creation itself. Instead, the interest lies in the underlying data of the artworks. From this perspective, databases would function merely as raw material for computational purposes.

purposes in educational establishments, without any profit intent; VII - the use of literary, artistic, or scientific works to produce judicial or administrative evidence; VIII - the reproduction in any work of small excerpts of pre-existing works of any nature or of entire works when it comes to plastic arts, provided that reproduction is not the main purpose of the new work and does not prejudice the normal exploitation of the reproduced work nor cause unjustified damage to the legitimate interests of the authors.

It is widely accepted that copyright protection concerns the original expression. Therefore, if the object of interest is considered to be limited to the data and not the work itself, it would be possible to classify the formation of databases for algorithm training solely as data mining, which is widely accepted as fair use.

The practice of data mining consists of recognizing patterns over a large amount of data based on large-scale computational analysis⁷. Even though the analyzed material is protected by copyright, the predominant understanding has been that this purpose would be covered by fair use. This is because the use of "text data mining" (TDM) allows for the discovery of new knowledge from the organization of existing knowledge. As it serves scientific and educational purposes, this has been the understanding of American courts regarding data mining (Krista L. Cox, 2015).

When confronted with the criteria proposed by fair use, the applicability of the mentioned doctrine for data mining cases is observed as follows: (i) regarding the purpose and character of the use, courts understand that the transformative nature of the use prevails, and it is demonstrated that the creation of a database differs in purpose, character, expression, and meaning from the material from which it is extracted; (ii) regarding the amount and substantiality of the portion used, it is understood that making copies of the entire work for analysis purposes was necessary; otherwise, it would be impossible to perform effective data mining, thus even if it involves substantial copying of the work, it is considered reasonable since the work is not used as an expression; and (iii) regarding the market impact, given its highly transformative nature, it is unlikely that using the work in databases will generate adverse impacts for the original from a market perspective.

When analyzing the nature of the copied material, it is observed that in cases where the analyzed data were underlying works with greater copyright protection, i.e., works with creative – and not functional – character, courts did not place much importance on this factor⁸, considering it neutral in most cases or moderately unfavorable to fair use applicability.

Additionally, to feed the algorithm training database, the application used the LAION dataset, a collection of data created by a German non-profit

⁷ The European Directive 2019/790 of April 17, 2019, defines it as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations;" (European Parliament, 2019).

⁸ This reasoning derives from the courts' understanding that the mere recontextualization of a protected work from an expressive context to another context is sufficient to apply fair use. According to Matthew Sag (2023), there were similar judgments in the U.S., such as in the case *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 609-610 (2d Cir. 2006), where the use of rock promotional posters to illustrate a biography had a different context from the original use for which it was created.

organization. LAION ("Large-scale Artificial Intelligence Open Network") is a database containing 5.85 billion images accompanied by texts, conceived by volunteers (The Batch, 2023) to democratize research (Beaumont, 2022).

In this sense, it seems reasonable to assume that the fair use doctrine would adhere to this modality, as the dataset's primary objective was to foster the development of machine learning model research in a public character.

Based on this reasoning, using images to create the database would fall under the fair use doctrine, considering the scientific-educational nature associated with the formation of this database. Applying AI would allow for the reuse of incorporated data, and the new works would promote scientific progress. However, a different analysis should be given when these data are used to train programs with profitable purposes.

b) Unfavorable Arguments for Applying Fair Use

Applying fair use to supply training data for generative AI is indeed quite controversial. Although databases such as LAION were created for research and technology development purposes, it is notable that platforms like Midjourney and LensAI use this technology to generate marketable digital goods. In other words, these platforms subvert the use of images to serve their profit-making purposes.

The LensAI application requires a monetary payment from the user for avatars to be generated. Similarly, Midjourney requires a subscription plan for the service to be provided. Thus, the purpose becomes purely commercial, which is a use not covered by the scope of fair use.

The main argument used to support the artists' plea refers to the market impact resulting from that use. According to a testimony given by artist Karla Ortiz when discussing AI and copyright before the U.S. Senate Judiciary Subcommittee on Intellectual Property:

[...] It is bad enough that this is being done without our consent, without any credit being offered, or without any compensation, but worse still, we are now forced to compete against these Generative AI models that were built upon our own work. No human being can outcompete a Generative AI model, due to the economies of scale: an AI is low cost and can generate a massive volume of "good enough" products especially compared to a single artist (United States of America, 2023).

Similarly, Samuelson (2023, p. 66) notes that these systems can produce results more quickly and economically than human authors, and AI results are generally of sufficient quality to be competitive in the market compared to human authors' works. This means that there is a non-trivial potential for market substitution. At this point lies the greatest controversy regarding using artworks to train algorithms. Generative AI

technology threatens artists' opportunities to compete in the market, causing a loss of income based on an exploitation structure of works whose use was neither consented to nor credited. There is not even any form of compensation, as noted by the litigants.

Regarding the market impact of applying fair use, it is also necessary to highlight the potential emergence of a market for creating specific licenses to authorize works to be included in such databases⁹. If the decision is to simply apply fair use, these licenses will become obsolete, causing significant losses to this market niche, which is another factor that should be weighed in deciding the case.

It is important to consider the separation between the database used and the generative AI system itself. However, if there is an infringement in the used database, Pamela Samuelson (2023b, p. 70) points out that courts may understand that the AI use is also "contaminated," which would automatically transfer the infringement to these systems.

Even though there is an evident separation between merely creating a database and the generative AI system, Pamela Samuelson suggests that U.S. courts might apply the same understanding to both hypotheses – as noted, the courts have been understanding fair use in data mining cases – so it will be up to the plaintiffs to "distinguish."

Finally, it is important to consider that there are solid arguments on both sides. However, it should be emphasized that the applicability of the fair use doctrine depends on the specific case analysis to weigh the interests involved in particular cases, which has not yet been analyzed by the courts.

V. CONCLUSION

The meaning and applicability of fair use for supplying digital art databases must consider not only public sphere interests for technology dissemination and development but also private sphere interests represented by the authors' property rights. The investigation demonstrated that it is possible to list arguments both for and against fair use application.

Due to the immediate functionality extraction of protected digital goods through simple sharing, the importance of regulating access protection was concluded, as this faculty is currently more relevant than the condition of effectively owning a particular good.

Throughout the investigation, the conflicting interests involved in using artworks in databases serving generative AI training were analyzed, which proved to be

⁹ Pamela Samuelson (2023b, p. 82) observes that the Copyright Clearance Center should explore the opportunity to organize licenses for the use of protected works by generative AIs, given that licensing markets have already emerged in various contexts of disruptive technologies. In this regard, the author points out that the judiciary has considered the threat to existing or potential markets, as in the case of the Napster and TVEyes applications.

a utilitarian issue. On one hand, there is the collective interest for technology to be improved and enhanced; on the other, there are the moral and property interests of the work creators.

To frame the debate, different viewpoints on the issue were discussed. One stance argues that the use does not incur any violation, as it would be analogous to the human inspiration process. According to this perspective, if the influence of others' works on the human author's creation is not considered misuse, then composing the database for algorithm training should be understood similarly. This perspective suggests there is no direct use of the contained arts but rather the processing and treatment of the data underlying them to define solutions by the algorithm.

In parallel, the artists' perspective argues that using works for machine learning is improper due to the similarity between AI products and the works contained in the database. In these terms, the necessity of artist authorization for such use was discussed in light of the argument that AI programs appropriate the databases compiling the works without crediting or compensating the authors.

In the end, the study compared different understandings of machine learning techniques. Some authors merely treat it as data mining, while others argue that it involves the exploitation of the work itself.

The cases analyzed highlighted that AI program developers seek to apply fair use in database formation by considering the use purpose and the work portion used. On the other hand, for the artists whose works are included as parameters, the market impact caused by using their work seems more relevant.

From the developers' perspective, the purpose of use to be considered is the development of machine learning techniques for technology progress, which would easily adhere to fair use. There are databases organized solely for research purposes, such as LAION; however, the study highlighted that some platforms use databases like this to provide new images that users pay for, thus subverting the research purpose compatible with fair use.

It is undeniable that producing new artistic works on a simple user command impacts the authors' income. Generative AI technology threatens artists' opportunities to compete in the market, causing income loss based on an exploitation structure of works whose use was not consented to and without remuneration for the artists. However, this argument is difficult to prove, as there is no way to precisely identify the copy of a specific original work in the AI result. In any case, it was concluded that fair use application requires specific case analysis, which has not yet been judged by the court where the action was filed.

Regarding the application of the fair use doctrine in Brazil, it was found more advantageous to

extensively interpret the limitations to copyright protection provided in Brazilian legislation, aiming to cover new hypotheses arising in the real world.

Although the claim of the authors, whose works are included in AI training databases, is legitimate for the recognition of their rights, it is technically difficult to prove these claims. This demonstrates that the ideal solution to the problem would be to regulate permission oversight by authorities. Such regulation would allow artists to consent to the use of their works and receive financial compensation for their valuable contribution to generative AI systems.

REFERENCES RÉFÉRENCES REFERENCIAS

- Alves, M. A. S.; Pontes, L. M. *O direito de autor como um direito de propriedade: um estudo histórico da origem do copyright e do droit d'auteur*. In: Congresso Nacionaldo Conpendi, XVIII. 2009, São Paulo. Disponível em: http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/sao_paulo/2535.pdf.p.9872
- Beaumont, R. Laion 5b: A new era of open large-scale Multi- Modal datasets. Laion. 31mar. 2022. Disponível em: <https://laion.ai/blog/laion-5b/>. Acesso em: 01 jun. 2024.
- Bentham, Jeremy. *Uma introdução aos princípios da moral e da legislação*. Tradução de Luiz João Baraúna. São Paulo. Editora Abril. 1974.
- Bobbio, Norberto. *Da estrutura à função: novos estudos de teoria do direito*. Tradução de Daniela Beccaccia Versiani. Barueri: Manole, 2007.
- Brasil. *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: SenadoFederal, 1988.
- Brasil. *Lei Nº 9.610*, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Brasília, DF
- Brasil. Senado Federal. *Projeto de Lei nº 5.691, de 05 de outubro de 2019*. Institui a Política Nacional de Inteligência Artificial. Brasília: Senado Federal, 2019. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=8031122&ts=1702407061693&disposition=inline&_gl=1*_ibzeks*_ga*NjYyMTA1ODAuMTY5NjlxMTU4OQ.*_ga_CW3ZH25XMK*MTcwNDY1OTlyMi41LjEuMTcwNDY1OTY5Mi4wLjAuMA.
- BRASIL. Superior Tribunal de Justiça. *Recurso Especial964.404/ES*. Relator Ministro PaulodeTarso Sanseverino, Terceira Turma, DJede23/5/2011.
- Brynjolfsson, E.; Clark, J.; Etchemendy, J.; Fattorini, L.; Ligett, K.; Lyons, T.; Manyika, J.; Maslej, N.; Ngo, H.; Niebles, J.; Parli, V.; Perrault, R.; Shoham, Y.; Wald, R. *"The AI Index 2023 Annual Report"* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.

10. Cox, Kristal. *Text and Data Mining and Fair Use in the United States*. Association of Research Libraries, 2015.
11. European Parliament news. *What is artificial intelligence and how is it used?* 4de setembro de 2020. Disponível em: <<https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used>>. Acesso em 22.05.2024.
12. Cerri, R.; Carvalho, A. C. P. L. F. Aprendizado de máquina: breve introdução e aplicações. *Cadernos de Ciência & Tecnologia*, v. 34, n. 3, p. 297-313, 2017.
13. Ehrhardt Jr., M.; Guilhermino, E. Breves notas sobre a (in) suficiência da teoria clássica da propriedade para disciplinar a titularidade dos bens digitais. In: Catalan, Marcos; Rocha, Mariângela; Pereira, Gustavo, orgs. *O caos no discurso jurídico: um homem a gema Ricardo Arrone*. Londrina. Editora Thoth, 2021.
14. Estados Unidos da América. *United States Copyright Act of 1976*. Public Law 94-553, 17 U.S.C. §§ 101 et seq., 1976. Disponível em <https://www.copyright.gov/fair-use/>. Acesso em: 22 mai. 2024.
15. Estados Unidos da América. *United States Copyright Act of 1976*. Public Law 94-553, 17 U.S.C. § 107, 1976, p. 20. Disponível em: <https://www.copyright.gov/title17/title17.pdf>. Acesso em: 22 mai. 2024.
16. Estados Unidos Da América. *United states copyright office*. Compendium of U.S. Copyright Office Practices. Third Edition. 2017.
17. Estados Unidos. *US Senate Judiciary Subcommittee on Intellectual Property "AI and Copyright"*. Written Testimony of Karla Ortiz. 2023. Disponível em https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_ortiz.pdf. Acesso em 31 mai. 2024.
18. Gil, A. C. *Como elaborar projetos de pesquisa*. 6. ed. Barueri: Atlas, 2017. Goodfellow, I. et al. Generative Adversarial Nets. *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*. Montreal, Canada, v. 2, p. 2672-2680, 2014.
19. Gpo. *National Artificial Intelligence Initiative Act of 2020*. Washington, DC, 2020.
20. Hurwitz, J.; Kirsch, D. *Machine Learning for Dummies*: IBM Limited Edition. Hoboken: John Wiley & Sons Inc., 2018. E-book.
22. Laion Blog. Disponível em <https://laion.ai/blog/laion-5b/>. Acesso em 08 de maio de 2024. Mill, J. S. *A Liberdade/Utilitarismo*. São Paulo: Martins Fontes. 2000.
23. Parlamento Europeu. Diretiva (Ue) 2019/790 Do Parlamento Europeu E Do Conselho de 17 de abril de 2019 relativa aos direitos de autor e direitos conexos no mercado único digital e que altera as Diretivas 96/9/CE e 2001/29/CE.
24. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/ALL/?uri=CELEX:32019L0790>. Acesso em 22 de maio de 2024.
25. Poli, L. M. *Direito Autoral: Parte Geral*. Belo Horizonte: Del Rey, 2008.
26. Sag, M. Copyright Safety for Generative AI, *Houston Law Review*, V. 61, n. 2. 2023.
27. Samuelson, P. Fair Use Defenses in Disruptive Technology Cases. *UCLA Law Review*. Nov. 2023.
28. *Generative AI meets copyright*. In: UC Berkeley Artificial Intelligence Research Group (BAIR). 2023. Disponível em: <https://www.youtube.com/watch?v=6sDGIrVO6mo>. Acesso em: 01 jun. 2024.
29. Schirru, L. *Direito Autoral e Inteligência Artificial: A Autoria e a Titularidade em Produtos da IA*. 2020. Tese. Doutorado em Políticas Públicas, Estratégia e Desenvolvimento. Instituto de Economia da Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 2020.
30. Senado Notícias. *Comissão de Juristas vai atualizar o Código Civil para revolução digital*. 24.08.2023. Disponível em <https://www12.senado.leg.br/noticias/materias/2023/08/24/comissao-de-juristas-vai-atualizar-codigo-civil-para-revolucao-digital>. Acesso em 15. dez. 2023.
31. Souza, A. R. de. As etapas iniciais da proteção jurídica dos direitos autorais no Brasil. *Justiça & História*. Vol 6. nº 11. 2006.
32. Taveira Júnior, F. T. *Proteção dos digital assets sob o enfoque dos direitos de personalidades*. Tese (Doutorado em Direito), Universidade de São Paulo. São Paulo, 2015, p. 49.
33. Tenório Filho, G. M. Freitas; Mallmann, Q. O fundamento filosófico do copyright à luz das concepções utilitaristas em Jeremy Bentham e John Stuart Mill. *Revista de Propriedade Intelectual-Direito Contemporâneo e Constituição*, v. 12, n. 1, p. 52-90, 2018.
34. The Story of Laion, the dataset behind Stable Diffusion. *The batch*. 2023. Disponível em <https://www.deeplearning.ai/the-batch/the-story-of-laion-the-dataset-behind-stable-diffusion/?ref=dl-staging-website.ghost.io>. Acesso em 09 de novembro de 2023.
35. Varella, P. *Como a Inteligência Artificial pode dominar o mercado de arte?* 2023. Disponível em: <https://arteref.com/inteligencia-artificial/como-inteligencia-artificial-pode-dominar-o-mercado-de-arte/>. Acesso em 30 mai. 2024.
36. We've filed a lawsuit challenging Stable Diffusion, a 21st century collage tool that violates the rights of artists. Because AI needs to be fair & ethical for everyone. *Stable Diffusion litigation*, 2023.

Disponível em: <https://stablediffusionlitigation.com/>.
Acesso em 02 de março de 2024.

37. Zygmunt, B. *Modernidade líquida*. Tradução de Plínio Dentzien. Rio de Janeiro: Jorge Zahar Editor, 2001.

