Nikola R. Milosavljević*

Original Scientific Paper
UDC: 347.78: 004.8
doi: https://doi.org/10.56461/SPZ_1118
Received: 30. 1. 2026.
Accepted: 4. 3. 2026.
Online First: 5. 3. 2026.

PROTECTING COPYRIGHT IN THE GENERATIVE ARITIFICIAL INTELLIGENCE ERA: *DISNEY* AND *UNIVERSAL* V. *MIDJOURNEY*

Summary

On June 11, 2025, Disney and Universal Studios filed a lawsuit against artificial intelligence (hereinafter: AI) company Midjourney for copyright infringement. This is the first lawsuit filed by major Hollywood companies against an AI company, and therefore it has the potential to establish a significant precedent for generative AI. Having this in mind, the author will attempt to examine the legal issues presented by generative AI and its implications for copyright protection. This paper is a comparative study of U.S. and European law with respect to the challenges posed by generative AI, and it proposes solutions to these issues, taking into account the differences in these legal traditions. Through the use of legal-dogmatic and comparative methods, as well as case study, analytic, and synthetic methods, the author aims to identify a universal solution to the global problem represented by AI – a problem that has been brought to the forefront by this and many other lawsuits.

Keywords: Generative AI, Copyright, AI Regulation, Midjourney Lawsuit, Copyright Infringement.

* PhD, Assistant professor, University of Kragujevac, Faculty of Law, Serbia.
E-mail: *nmilosavljevic@jura.kg.ac.rs*
ORCID: *https://orcid.org/0000-0003-0208-2807*

ZAŠTITA AUTORSKIH PRAVA U ERI GENERATIVNE VEŠTAČKE INTELIGENCIJE: PREDMET *DISNEY* I *UNIVERSAL* PROTIV *MIDJOURNEY*

Sažetak

Dana 11. juna 2025. godine, Disney i Universal studio podneli su tužbu protiv kompanije za veštačku inteligenciju (VI) Midjourney zbog kršenja autorskih prava. Ovo je prva tužba jedne velike holivudske kompanije protiv kompanije za veštačku inteligenciju i mogla bi da postane važan presedan za regulisanje generativne VI. Imajući to u vidu, autor će pokušati da ispita pravni problem generativne VI i njegove posledice po zaštitu autorskih prava. Rad predstavlja uporednu studiju prava SAD i Evropske unije u vezi sa problemom generativne VI i nudi moguća rešenja ovog problema, imajući u vidu razlike u pomenutim pravnim tradicijama. Primenom pravnog dogmatskog metoda, uporednopravnog metoda, studije slučaja, analitičkog i sintetičkog metoda, autor će pokušati da nađe univerzalno rešenje za globalni problem generativne VI, koji je sada aktuelizovan ovim i mnogim drugim slučajevima.

Ključne reči: generativna VI, autorsko pravo, regulisanje VI, tužba protiv Midjourney, povreda autorskih prava.

1. Introduction

On June 11, 2025, Disney and Universal filed a lawsuit against AI startup Midjourney in a federal district court in Los Angeles for copyright infringement. Although there have already been copyright infringement cases brought against AI companies, this marks the first instance in which two major Hollywood companies have entered such a legal dispute. Therefore, this case is likely to attract significant public attention. A decision in this case could potentially alter prevailing understanding of copyright law and, consequently, bring about lasting changes in the entertainment industry. In this paper, we will examine the known facts of the *Disney and Universal v. Midjourney* case as reported in the press, and conduct a review of earlier copyright infringement cases involving AI. We will then undertake a comparative analysis of U.S. copyright law and the copyright laws of European countries with respect to generative AI copyright infringement. Finally, we will draw conclusions from the conducted analysis and propose possible solutions to the problem of generative AI copyright infringement.

2. Facts of the *Disney and Universal v. Midjourney* Case and Preceding Case Law

According to press reports, Midjourney enabled subscribers to generate images using its AI system. The AI system produced content in accordance with prompts provided by subscribers, utilizing images and videos from the internet (Knibbs, 2025). In this way, Midjourney reportedly generated USD 300 million in revenue in 2024 (Chmielewski, 2025). Many of the prompts included well-known characters such as *Darth Vader* and *Yoda* from *Star Wars*, *Iron Man, Shrek, Spiderman,* and others. These characters originate from films or animated works for which Disney and Universal hold copyright. Consequently, the studios assert that Midjourney, as a provider of generative AI services, infringed their copyrights and are seeking damages of USD 150,000 for each copyrighted work, in addition to an injunction requiring Midjourney to cease further copyright infringement (Duffy, 2025). The plaintiffs' legal team stated in the press that, while AI plays an important role in technological development and has benefited the entertainment industry, piracy remains illegal regardless of whether it is committed by humans or by AI (Montgomery, 2025). They further characterized Midjourney as "a bottomless pit of plagiarism" (Duffy, 2025). The plaintiffs also asserted that, although they do not object to Midjourney's use of their copyrighted works, the company must compensate them for such use (Chow, 2025). At the time of writing, Midjourney has not yet responded to the lawsuit; however, its management stated at a conference that they were not concerned about the company's future (Chen & O'Brien, 2025).

Previous cases were similar to the present one, such as the dispute between *The New York Times and ChatGPT* (Montgomery, 2025; Mešević, 2024). Another notable example is *Authors Guild v. Open AI,* a copyright infringement case in which ChatGPT was alleged to have been trained on copyrighted texts and books (Culliton, 2024, p. 263). The *Doe v. GitHub* case involves anonymous programmers who claim that GitHub uses their code without their permission to develop its AI model (Napitupulu et al., 2023). Across the United States, numerous lawsuits have been filed against generative AI companies by authors, composers, and other artists and copyright owners seeking to protect their works from AI exploitation. Some of these cases have resulted in decisions favoring the copyright owners, while others have favored the generative AI companies (Knibbs, 2025). One of the most prominent cases involving generative AI is *Getty Images v. Stability AI*, initiated in February 2023. Getty, a company that maintains a substantial collection of images, argues that Stability AI infringed its copyrights by training the Stable Diffusion AI model[1] on its images. Getty sued in both the UK and the United States (Garon, 2023, p. 186).[2]

[1] Midjourney also used a model based on Stable Diffusion (Sobel, 2024, p. 55).

[2] Interestingly, although China has often been perceived as a rogue that violates intellectual

3. Comparative Analysis of U.S. and European Law Regarding Generative AI

3.1. U.S. Law

First, it is necessary to examine how generative AI operates. Generative AI represents a subfield of deep learning focused on generating outputs in response to user prompts or inputs in the form of commands (Napitupulu *et al*., 2023, p. 275). There are three types of generative AI: 1) Generative Adversarial Networks (GAN), which specialize in generating visual or multimedia outputs using both language and image inputs; 2) transformer-based models, such as Generative Pre-trained Transformers (GPT), which are designed for generating textual content; and 3) variational autoencoders, which use encoders to extract the underlying structure of the data in order to learn how it is constructed, and subsequently use it to generate images, text, or videos (Apoorva, 2023, p. 72). Midjourney and Stable Diffusion operate on the principle of autoencoding (Sag, 2023, p. 317), also referred to as a diffusion model (Sobel, 2024, p. 55). Initially, images, and other copyrighted materials are crawled from the internet by the company Common Crawl. These materials are then encoded in the word-image pairs and represented numerically, a process known as tokenization (Widła, 2025, p. 5). In other words, inputs and outputs are not processed merely as strings of words or pixels but as vectors or abstract representations (Sobel, 2024, p. 61). The crawled material is subsequently filtered to remove inappropriate content, such as pornography, as well as double entries, resulting in a foundation model that is used for AI training. A foundational model is a machine learning model trained on large-scale, broad datasets – typically using self-supervised learning – and is not trained to accomplish a specific task, but rather to capture generalizable and useful information from the data (Henderson *et al*., 2023, p. 4). One of the most widely recognized foundation models is LAION-5B, which has been used to train Stable Diffusion, DALLE-E, and ChatGPT (Murray, 2023, p. 285). The AI models are subsequently further trained to perform specific tasks, such as image generation. The AI program learns to create new content by analyzing existing content and predicting, based on that analysis, which element – whether a word or a pixel – should come next. Ultimately, it decodes the encoded content (i.e., image-word pairs) through a process of diffusion and denoising, using the results to generate new content based on its training (Murray, 2023, p. 295). This process is somewhat problematic for jurists. It would be inaccurate to assert that AI simply copies the copyrighted content and reassembles it in the generated output. Rather, AI utilizes copyrighted content to learn how to create new content (Murray,

property rights, the court in Guangzhou ruled that an AI company infringed copyright by using copyrighted works without authorization (Lu, 2024, p. 176).

2023, p. 286). The encoding and subsequent decoding of copyrighted material typically breaks the connection between the original expression in the copyrighted material and pseudo-expression manifested in the generated content (Sag, 2023, p. 318). It is not uncommon for AI to "memorize" content on which it has been trained and subsequently reproduce it in its entirety (Sag, 2023, pp. 313-321). This phenomenon, referred to as "memorization," is considered a common bug in AI programs and is reported to occur less frequently when the model is trained on larger datasets (Henderson *et al*., 2023, p. 17). Nevertheless, memorization can be triggered by a user request for the generation of copyrighted material (Cyphert, 2024, p. 53), and studies suggest that such occurrences are not rare (Sag, 2023, p. 327). Typically, AI reproduces only portions of the original work; however, in some instances, it may memorize even the entire work (Sag, 2023, p. 336). As evidenced by Disney's claim, AI can memorize copyrighted characters such as Shrek and others, which formed the basis of the lawsuit. When AI memorizes copyrighted characters, the argument for copyright infringement is particularly strong. This principle was also affirmed in the judgment of the Munich Regional Court (*Landsgericht München*) in *GEMA vs. Open AI* (2025, paras. 200-205), where it was determined that memorization occurs frequently and should be regarded as reproduction, constituting copyright infringement. Still, not all characters fulfill the criteria for copyright protection, as copyright generally protects the work as a whole.[3] Moreover, any reproduction must be substantially similar to the copyrighted work to qualify as infringement.

Apart from the aforementioned cases involving memorization of copyrighted characters, instances of substantial similarity between AI-generated content and copyrighted works are expected to be rare. In *Andresen v. Stability AI*, on October 30, 2023, several claims against the defendant were dismissed, including the allegation that AI-generated works constitute derivative works.[4] However, the court allowed the claim that copyright was infringed by training the AI on copyrighted works to proceed (Gendron, 2024, p. 180). A similar outcome occurred in *New York*

[3] According to the precedent established in the *Batmobile* case, a character from a work of authorship (e.g., a book, film or other creative works) is eligible for copyright protection if it meets the following requirements: (1) the character possesses both "physical as well as conceptual qualities"; (2) the character is "sufficiently delineated to be recognizable as the same character whenever it appears" and "display[s] consistent, identifiable character traits and attributes"; and, (3) the character is "especially distinctive" and "contain[s] some unique elements of expression." (Sobel, 2024, pp. 89-90).

[4] With the exception of Chinese law (Lu, 2024, p. 177), content created by AI is not eligible for copyright protection (Apoorva, 2023, p. 79). Therefore, classifying such content as a derivative work is also out of question. Derivative works must incorporate elements of the original work; however, this occurs only sporadically, as copyrighted material is used primarily for training purposes (LaPorta, 2024, p. 80; Gendron, 2024, p. 179).

Times v. Open AI (Cyphert, 2024, p. 62). This issue is interesting because, under Section 106 of Title 17 of the U.S. Code (U.S. Code),[5] the use of the copyrighted material for AI training, or for text and data mining (hereinafter: TDM), is not provided as an exclusive right of the copyright owner (Sobel, 2024, p. 60). Nevertheless, it would not be unreasonable to argue that authors should hold an exclusive right in this context. AI developers have argued that without copyrighted works, the development of generative AI would not be possible (Cyphert, 2024, p. 52). This essentially implies that AI companies utilize copyrighted material without authorization to develop products that may compete in the market with the copyrighted materials that was used for their training. Still, for such use to constitute infringement, text and data mining (TDM) would need to be recognized as an exclusive right. Some argue that TDM is a form of reproduction, as the method or technique used for reproduction is immaterial. Others contend that TDM is analogous to copying the source code of a computer program; even if the audio and visual elements of the program are not directly reproduced, copying the source code ultimately results in the reproduction of the visual elements. (Murray, 2023, p. 280; Sobel, 2024, pp. 65-68). However, the High Court of England and Wales rejected this argument in its judgement in *Getty Images vs. Stability AI* (2025, para. 600), holding that AI model does not constitute an infringing copy. Some scholars also maintain that most AI models do not perform actual copying (Sag, 2024, p. 1894). It is an established fact that images are encoded and, in the encoded form, transferred into a database. This process is more similar to reading than to classic copying because the word-image pair (token) created during crawling is not the image itself but the code used to train the AI. It could be argued that this represents copying the image in an abstract form; however, it cannot be used in that form for human consumption, nor is it fully replicated after decoding. Hence, it is not quite a forward line of reasoning that that text and data mining (TDM) constitutes reproduction. (LaPorta, 2024, pp. 83-84).

Most AI companies, when named as defendants, have relied on the fair use defense against claims brought by copyright owners (Dutta, 2025, p. 7769). Fair use is a doctrine specific to common law jurisdictions, allowing the use of copyrighted

[5] Section 106 of Title 17 of the U.S. Code provides the following exclusive rights: 1) to reproduce the copyrighted work in copies or phonorecords; 2) to prepare derivative works based upon the copyrighted work; 3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending; 4) in the case of literary, musical, dramatic, and choreographic works, pantomimes, motion pictures, and other audiovisual works, to perform the copyrighted work publicly; 5) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and 6) in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission.

material without authorization or payment under certain justified circumstances (Kuker, 2024, p. 5). In the United States, this doctrine is codified in Section 107 of Title 17 of the U.S. Code (U.S. Code). It provides that, in exceptional cases, a court may permit unauthorized use of copyrighted content, provided that it took into account the following four criteria: (1) the purpose and character of the use, including whether such use is commercial or for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work. The interpretation of these criteria is often interrelated, and judges and scholars do not always agree on their precise meaning (Lin, 2024, p. 477). Each factor plays a significant role, although the interaction among them is not always clear. Consequently, the fair use doctrine is surrounded by a degree of legal uncertainty, as arguments can be made both in favor and against fair use in any given case (Henderson *et a*l., 2023, pp. 4-5).

The purpose and character of the use is generally understood as a criterion related to how the copyrighted work will be utilized. The main question is whether the new work serves as a substitute for the copyrighted work, i.e., whether it has the same purpose and objectives (Culliton, 2024, p. 269). Typically, if the work is used for purposes of greater public importance, such as education or the advancement of science or culture, it is more likely to qualify as fair use. Studies show that in 77% of cases where courts upheld fair use, the use was non-commercial (Henderson et al., 2023, p. 7). On the other hand, commercial use provides fewer grounds for fair use, as copyright law also protects commercial interests (Kuker, 2024, p. 9). There are cases when the use of copyrighted works for AI training will be allowed, such as AI development in scientific research; however, it is essential to determine whether any commercial use is involved. A lack of direct sales will not suffice as an argument that there is no commercial use, as AI companies often charge subscription or usage fees for their programs (Garon, 2023, p. 190). Nonetheless, most AI companies have invoked a transformative use defense, a recognized form of fair use (Gendron, 2024, p. 168). They argue that AI programs utilize copyrighted materials to generate new content, which is generally permissible as it contributes to cultural development. For transformative use to be recognized, the new work must add something to the existing corpus, convey a different message, or produce a new meaning or expression, as established in *Campbell v. Acuff-Rose Music*. The degree of transformation depends on the context of each case (Henderson *et a*l., 2023, p. 6). A transformative work must supersede the objects of original creation in the initial work. Transformative use is closely linked to public use, as it promotes free speech and facilitates the creation of new, original expression (Murray, 2023, p. 272; Lin, 2024, p. 476). However, its applicability to AI-generated content is somewhat

debatable, since the transformation and development are not performed by humans and are not directly connected to the human need for self-expression (Lin, 2024, p. 491). Conversely, because copyrighted works are used solely as learning patterns, this process could arguably constitute transformative use (Culliton, 2024, p. 270). Some authors argue that AI use qualifies as transformative because text and data are transformed into numerical code (Murray, 2023, p. 280). It can also be argued that AI may be used for purposes other than copyright infringement, and therefore, such use could be considered fair (Murray, 2023, p. 311).

The nature of the copyrighted work is a criterion concerned with the qualities of the work itself. The more original the copyrighted work is, the less likely its use will be deemed fair (Kuker, 2024, p. 10). Courts are tasked with examining whether a copyrighted work primarily consists of original material or of elements that are not copyrightable or attributable to the author (Henderson *et al.*, 2023, p. 7). The scope of copyright protection is defined by the author's original expression. On the other hand, facts, ideas, methods, techniques, and theories are not protected by copyright. For this reason, non-expressive uses of copyrighted material are generally more likely to qualify as fair use (Sag, 2024, pp. 1900-1902).[6]

The amount of the portion used is another important criterion. The main question is whether the amount of the portion used is reasonable under the circumstances (Kuker, 2024, p. 10). Use of a smaller portion generally supports a finding of fair use, whereas extensive use tends to weigh against it. The sustainability of the use is also important; the portion used must not effectively replace the entire content of the copyrighted work (Sag, 2024, pp. 1902-1906). It is not only the quantity but also the qualitative value of the portion that is important; copying "the heart of the work," which embodies the most elements, weighs against fair use (Culliton, 2024, p. 273). If only the functional, rather than expressive, elements of a work are copied, even a substantial portion may be used (Henderson et al., 2023, p. 16). Conversely, copying the original and key elements of a copyrighted work supports a determination against fair use. Having in mind that AI systems typically utilize smaller portions of copyrighted works and generally avoid the key elements, this criterion would likely favor a finding of fair use. However, when AI memorizes copyrighted characters, this factor weighs against the application of the fair use doctrine. Fan

[6] One of the most interesting of elements that is generally not copyrightable is style. Style refers to the distinctive manner of expression of a particular artist, and, like a method, it cannot be copyrighted. However, it is important to note that when style represents a part of the author's expression, it may qualify for copyright protection. Distinguishing between personal expression and elements that do not constitute copyrightable expression has traditionally been challenging. Until now, an author's personal style could never be fully reproduced, so this was not a concern. However, generative AI can reproduce an author's personal style almost entirely, raising new questions regarding the copyrightability of personal style (Sobel, 2024, p. 72 et seq.).

fiction is generally not accepted as fair use; therefore, creating stories or videos based on copyrighted characters using AI would likely constitute infringement (Henderson *et al*., 2023, p. 12).

The effect of the use on the potential market is likely to be the most interesting factor for courts to assess in generative AI-related cases. Scholars believe that this factor will experience a renewed prominence in the context of generative AI litigation (Lin, 2024, p. 492). Courts must take public policy considerations into account, including the potential negative effects that the overflow of AI-generated works may have on human creativity and innovation (Culliton, 2024, p. 282). AI-generated works can easily flood the market with inexpensive works of quality comparable to that of copyrighted works. One of the main reasons copyright owners initiate litigation is the concern that AI-generated works may easily substitute their original creations, thereby depriving them of income. Because AI-generated works can be created fast and at low cost, they pose serious competition to copyrighted works that require significant time and effort to be created (Culliton, 2024, p. 274). Therefore, this factor is likely to weigh against a finding of fair use. However, economic damage to copyright owners cannot be presumed and must be demonstrated, particularly given that the quality of AI-generated works is often not equivalent to that of copyrighted works (Kuker, 2024, pp. 23, 29). In connection with the first fair use factor, and in light of *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, courts must also consider the new market function of generative AI and how it affects copyright owners (Lin, 2024, p. 489). This judgment implies that courts, when determining whether fair use should be permitted, are required to perform a cost-benefit analysis.[7]

3.2. The EU Law

Most scholars agree that the European Union has already regulated the relationship between copyright law and generative AI through the text and data mining (TDM) exceptions provided in Articles 3 and 4 of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (hereinafter: the CDSM Directive) (Dornis, 2025, p. 66). This interpretation was also adopted by the Hamburg Regional Court (*Landsgericht Hamburg*) in case no. 310 O 22723 of September

[7] If courts permit fair use in this context, this may lead to several adverse consequences, including the development of deepfakes and misinformation, the disclosure of private information, the reinforcement of biases, cultural homogenization and conformity, an unhealthy dependence on technology, and a chilling effect on the cultural sector. Conversely, if courts decline to allow fair use, this may also produce significant negative effects, such as reduced productivity, impediments to technological development and emerging forms of creativity, limitations on cost reductions in education, and a chilling effect on scientific research (Sag, 2024, pp. 1896-1898).

27, 2024 (LAION) (Dusollier *et al*., 2025, p. 125), as well as of the Munich Regional Court in the aforementioned *GEMA vs. OpenAI* (2025, para. 243), although the latter court ultimately did not apply the exception in that particular case. Article 3 CDSM Directive provides that Member States shall establish an exception to the rights of reproduction and extraction for research organizations and cultural heritage institutions, permitting them to carry out TDM for the purposes of scientific research, provided that they have lawful access to the works or other protected subject matter (CDSM Directive). This exception has been frequently criticized as being overtly narrow for the purposes of generative AI development. First, this exception applies only to the right of reproduction as a core right of copyright and extraction in the case of databases. Moreover, reproduction and extraction are permitted only when TDM is performed by research organizations or cultural heritage institutions and solely for the purposes of scientific research (Margoni & Kretschmer, 2022, p. 694). Consequently, both conditions have to be satisfied, which significantly narrows the scope of application (Kollár, 2022, p. 18) and renders the exception poorly suited for cooperation between research institutions and the private sector (Dermawan, 2023, p. 53). It is also crucial that the beneficiaries operate on a nonprofit basis, as the exception would not apply otherwise (Moerland, 2024, p. 679; Ivanović, 2021, p. 70). Furthermore, this exception does not apply to computer programs. The exclusion of computer programs from the scope of the exception is somewhat surprising and unclear; however, it is assumed that this limitation was introduced due to concerns related to potential reverse engineering (Widła, 2025, pp. 8-9). The most significant criticism of this provision, however, lies in the requirement that research organizations must have lawful access to the protected content. Lawful access generally means that the right holder has made the work or database publicly available, or that the research organization or cultural heritage institution has obtained a license or other authorization for accessing the protected works. Because access licenses are subject to the discretion of the right holder, the right holder can effectively disable this exception (Szkalej, 2025, p. 317). Nevertheless, it is a general principle applicable to all copyright exceptions that beneficiaries must have legitimate access to the protected works, meaning that they may not use pirated versions of the work (Szkalej, 2025, p. 312). Moreover, lawful access in this context simply means that beneficiaries may perform TDM if they can read or otherwise consume the work. No specific license for TDM is required, as TDM is not recognized as an exclusive right of the copyright owner (Moerland, 2024, p. 678; Bottis *et al*., 2019, p. 195), although some argue that the absence of an explicit TDM license requirement is not entirely clear from the wording of the provision (Rosati, 2019, pp. 214-215).[8] The requirement of lawful access is intended to ensure that right

[8] This does not imply that the right holder cannot make TDM difficult or even impossible by using technical means; therefore, caution is warranted (Szkalej, 2025, p. 315).

holders are adequately compensated (Kollár, 2022, p. 15; Ivanović, 2021, p. 71). Nevertheless, scholars have correctly observed that this requirement may lead to increased licensing costs for ordinary consumption of works, as right holders may incorporate the value of TDM license into their overall price (Kollár, 2022, p. 17).

Article 4 CDSM Directive introduces an exception that is available to all potential beneficiaries, including commercial entities. It provides that Member States shall establish an exception or limitation to the rights of reproduction and extraction of lawfully accessible works and other subject matter for the purposes of text and data mining (CDSM Directive). This exception does not specify particular beneficiaries or purposes, making it broader in scope than Article 3. Unlike Article 3, it also applies to computer programs. However, paragraph 3 of Article 4 stipulates that the exception or limitation is subject to the condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by the right holders in an appropriate manner, such as through machine-readable means in the case of content made publicly available online (CDSM Directive). This effectively means that the Article 4 exception can be disabled at the discretion of the right holders if they choose to "opt out" their content from TDM, thereby unilaterally preventing TDM – a practice that has been widely criticized (Kollár, 2022, p. 19). The ability to opt out of an exception is relatively unusual in classical exception theory. Traditionally, the core principle underlying exceptions in intellectual property law is that there exist circumstances in which the public may use the protected intellectual property without the consent of the right holder. Under this framework, the responsibility for opting out lies with the right holder, and not with the beneficiaries. As noted in paragraph 3, right holders must opt out expressly, which, in the case of publicly available content on the internet, requires the use of technical means to prevent their content from being crawled (Moerland, 2024, p. 699; Dermawan, 2023, p. 53). This could include, for example, robot.txt files or meta tags, although any other machine-readable option is also permitted (Lobling *et al*., 2023, pp. 505-506). Nevertheless, this does not preclude the possibility of *ex post* opt-out, since under the recently enacted EU Artificial Intelligence Act (EU AIA), developers of large language models are required to transparently make available all content on which their LLMs were trained, thereby enabling right holders to protect their rights and potentially exercise an opt-out (Dutta, 2025, p. 7772). Nevertheless, in cases where an AI model has already been trained on copyrighted content, it is not possible to reverse the training process, making an *ex ante* approach more reasonable (Popović, 2024, p. 48). For content that is not publicly available, it is sufficient for right holders to include an opt-out declaration in the terms and conditions (Lobling *et al*., 2023, p. 504), or to employ a *Creative Commons* license that explicitly excludes TDM (Guadamuz & Cabell, 2014, p. 19). In all cases, lawful access remains a fundamental requirement.

Regulating text and data mining (TDM) in this manner has been widely criticized by scholars as not conducive to a friendly AI policy (Margoni & Kretschmer, 2022, p. 686; Dermawan, 2023, p. 52; Ivanović, 2021, p. 73; Rosati, 2019, p. 209; Perišić, 2025, p. 620). The fact that these exceptions are limited to the rights of reproduction and extraction, that the scope of Article 3 is very narrow, that right holders may opt out of the Article 4 exception, and that in both cases beneficiaries must have lawful access to the protected content, collectively make it extremely difficult for AI developers to perform TDM under these exceptions. This regulatory framework disproportionately affects small and medium-sized enterprises, placing them at a disadvantage compared to large corporations that already have access to copyrighted material or can obtain it through their financial resources (Margoni & Kretschmer, 2022, p. 687; Dermawan, 2023, p. 58). However, as rightly noted in the literature, these exceptions apply only when TDM includes the copying of copyrighted material or the extraction of a database. TDM does not necessarily include copying and extraction because, as discussed earlier, copyrighted material is encoded without reproducing it in its original form. Therefore, since TDM is not an exclusive right of the author or database owner, it is fully permissible if no act of reproduction or extraction occurs[9] (Orlando & Casalini, 2023, p. 7; Radovanović, 2018). It is also important to note that some scholars argue that TDM inherently involves some form of reproduction or extraction, regardless of how the data is transferred, and thus these exceptions would always apply (Moerland, 2024, p. 676; Ivanović, 2021, p. 75). This position is generally accepted in the case for computer programs (Widła, 2025, p. 7). It is highly likely that the act of TDM would not fall within the notion of reproduction under Article 9 of the Berne Convention, since no recognizable copy of the work exists after TDM is performed, and the copyrighted work is not used "as work." However, the concept of reproduction is not universally defined, and each state may adopt its own interpretation (Orlando & Casalini, 2023, p. 14; Senftleben, 2022, p. 1494; Ivanović, 2021, p. 65). As noted previously, this interpretation was reaffirmed by the High Court of England and Wales in the judgment of *Getty Images vs. Stability AI* (2025, para 600).

In our view, the main question is not whether text and data mining (TDM) includes reproduction or extraction, but rather what TDM represents in the generative AI process. According to Article 2 CDSM Directive, TDM means any automated analytical technique aimed at analyzing text and data in digital form in order to generate information, including, but is not limited to, patterns, trends, and correlations (CDSM Directive). At the time when the exceptions were proposed (2016)

[9] This is typically the case in scientific research, as the data on which the research is performed must be preserved both as evidence and to ensure the reproducibility of the research (Margoni & Kretschmer, 2022, p. 694).

and when the CDSM Directive was enacted (2019), TDM was understood quite differently from the auto-encoding process used for AI training. Specifically, TDM was regarded as an activity involving the analysis of vast amounts of data and/or text to extract knowledge or information (Kollár, 2022, p. 3; Moerland, 2024, p. 673; Ivanović, 2021, p. 61). It was believed that AI would facilitate data analytics; instead of manually reading and classifying vast amounts of data to draw conclusions, AI could perform TDM and extract the necessary information within minutes (Guadamuz & Cabell, 2014, p. 5). As evident, the material generated through TDM should consist of information such as patterns, trends, or correlations, rather than quasi-copyrighted content (Moerland, 2024, p. 677). This aligns with the perception that AI, when performing TDM, does not copy the author's original expression, but rather extracts information, facts, and knowledge embedded in the data and text, which are not protected (Margoni & Kretschmer, 2022, p. 698; Dermawan, 2023, p. 49). Therefore, since AI does not extract the original expression but rather information and facts, which are not protected by copyright, there can be no exclusive right of the author over TDM (Margoni & Kretschmer, 2022, p. 688). However, as discussed above, during training, generative AI extracts more than just facts, information, and knowledge; it also captures the author's own expression in sentences, drawings, or even copyrighted characters. In other words, AI mines not only the semantics of the work but also its syntax (Dornis, 2025, p. 68). Consequently, it would be incorrect to equate TDM performed in the course of training of generative AI for creating images and videos with TDM performed by research institutions or marketing organizations for the purpose of analyzing collected data and extracting information (Dornis, 2025, p. 67). This view is also reflected in the European Commission's *Study on copyright and new technologies: copyright data management and artificial intelligence* (European Commission, Directorate General for Communication, 2022, p. 205). The fact that Article 53 of the EU Artificial Intelligence Act (EU AIA) refers to Articles 3 and 4 of the CDSM Directive does not imply that TDM constitutes an exception regulating generative AI, as is often misunderstood (Szkalej, 2025, p. 311). This is because classical TDM may also be performed by AI, but it does not necessarily involve generative AI as a specific type (Dornis, 2025, p. 71). However, another view upheld by the European copyright community is that the TDM exceptions cover a part of the generative AI process, though not its entirety, specifically the crawling component. This view was also adopted by the Hamburg Regional Court in the *LAION* ruling (Dusollier *et al*., 2025, pp. 122-123) and by the Munich Regional Court in *GEMA vs. OpenAI*. Their position is that the definition of TDM is sufficiently broad to encompass the generation of new works, as these works constitute new information.[10] However, we contend that this

[10] It is fair to say that Munich Regional Court did not apply these exceptions in its case because

interpretation is excessively broad for the reasons discussed above. The underlying purpose of the TDM exceptions was to permit the analysis of text and data using AI in order to acquire knowledge or useful information, and not to generate content resembling the analyzed works. Equating works of authorship with information generated from data analysis would be an oversimplification. The significance and potential impact on authors are far greater than that of generative AI, particularly considering that not only knowledge but also the author's original expression is extracted – and subsequently imitated – by AI. For these reasons, we maintain that the exceptions provided in Articles 3 and 4 of the CDSM Directive do not extend to generative AI, as applying them in this context would constitute an excessive limitation of the author's exclusive rights and would fail to satisfy the three-step test established under the Berne Convention. The most problematic aspect is the second step of the three-step test – namely, that the exception must not interfere with normal exploitation of the work.[11] This condition is not met if the exploitation of the work under the exception enters into economic competition with the ways in which right holders normally derive economic value from their works. While this is undoubtedly true for TDM that is provided under Articles 3 and 4 of CDSM Directive (Kollár, 2022, p. 27), it is unlikely to hold true for TDM performed in the context of generative AI training, as the resulting products – literary and artistic works – directly compete with the copyrighted content used during training.

4. Possible Solutions to the Dispute and Approaches for Regulating Generative AI

The solution of *Disney v. Midjourney*, *Getty Images v. Stability AI* in the U.S. (and the first-instance judgment in the UK case), as well as other generative AI cases lies ahead of us, and it may take several years before the final solution is reached. It remains uncertain how the courts will decide in these cases. They will first need to determine whether the TDM performed during crawling constitutes reproduction of the copyrighted material, although, as noted earlier, the High Court of England and Wales did not recognize it as such. The question of the fair use doctrine will require courts to examine whether the use of copyrighted material for AI training is transformative and whether the technical development of AI is important enough to outweigh copyright protection. Conversely, they will need to consider whether

the AI model performed memorization, which, in the court's view, could not be covered by the exceptions.

[11] This position is also recognized by the European Copyright Society (Dusollier, *et al*., 2025, p. 123) and supported elsewhere in the literature (Popović, 2024, p. 47).

the economic impact of generative AI on copyrighted works is so severe that fair use cannot be justified. In this context, courts may also look to align their reasoning with European law, reflecting a broader trend of aligning the EU and the U.S. practices. They may assess whether the requirements for the exception under Article 4 CSDM Directive are met and align their approach with that of the regional courts in Hamburg and Munich. For example, Stable Diffusion provides authors with an opt-out option, which it cites as part of its defense (Sobel, 2024, p. 61).

We are of the opinion that TDM should be recognized as an exclusive right of the author. There are several reasons supporting this view. First, TDM cannot, in most cases, be equated with copying or reproduction of the work. Even when it involves some form of reproduction, the purpose is not to present the work but to train AI, which is fundamentally different from the classical notion of reproduction. Moreover, as discussed earlier, TDM in the context of generative AI goes beyond filtering information and knowledge from copyrighted content; it involves using the entirety of the copyrighted work for training, including the author's original expression (Dornis, 2025, p. 61). Hence, AI training represents a form of using original parts of an author's work, for which the author has a right to be compensated.[12] Although this constitutes a form of non-expressive use, the fact remains that AI leverages copyrighted works to create similar content that competes in the market for artistic works. Generative AI can, for a relatively small membership fee, effectively substitute works created by human authors. For all those reasons, we believe that a special right on TDM should be established. However, generative AI has proven to be highly valuable and important for technical advancement, and it is necessary to create conditions that allow its development (Živković, 2024, p. 319), even though it also brings certain negative consequences (Avramović & Jovanov, 2023, p. 163). A balance between these two important interests could be achieved by introducing a statutory license for copyrighted works. Under such system, AI companies would be allowed to use copyrighted works for training without needing the author's individual permission, provided that they pay remuneration to the copyright owner. This approach could prevent copyright owners from inflating royalties in a way that might chill AI development, while simultaneously ensuring that authors are compensated for the use of their works. This still means that some form of collective protection would likely be necessary to facilitate the collection and distribution of payments to copyright owners.

[12] The necessity of remuneration for TDM conducted in the context of generative AI has also been recognized by the European Copyright Society (Dusollier et al., 2025, p. 126).

5. Conclusion

Disney and Universal v. Midjourney represents yet another confrontation between copyright owners and generative AI companies. Unlike previous cases, this is the first instance in which such major corporations have pressed charges against an AI company, signaling that generative AI has become a significant challenge for copyright protection and copyright owners' interests. Analyzing both U.S. and EU law, it becomes clear that neither system is fully adapted to address the complexities of generative AI. U.S. law is arguably in a somewhat better position due to the flexibility of its precedent-based system and the fair use doctrine, which allows courts to tune and adjust the law to respond to generative AI development. EU law is not without potential solutions. However, in our view, Articles 3 and 4 of CDSM Directive cannot be directly applied to TDM performed for generative AI training. Nevertheless, they could serve as a starting point for developing a framework that would balance the interests of both AI developers and the creative industry. In our opinion, the most effective solution would be the introduction of a statutory license, which would protect both the economic interests of copyright owners and at the same time prevent a chilling effect on AI development – an activity that has been recognized as one of the most important public interests of our time.

References

Apoorva. V. 2023. The copyright problem with emerging generative AI. *Journal of Intellectual Property Studies*, 7(2), pp. 69-84.

Avramović, D. S. & Jovanov, I. D. 2023. Sudijska (ne)pristrasnost i veštačka inteligencija. *Strani pravni život,* 67(2), pp. 161-177. https://doi.org/10.56461/SPZ_23201KJ

Bottis, M., Papadopoulos, M., Zampakolas, C. & Ganatsiou, P. 2019. Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit. *Erasmus Law Review*, 12(2), pp. 190-208. https://doi.org/10.5553/ELR.000140

Chen, S. & O'Brien, M. 2025. Disney and Universal sue AI firm Midjourney for copyright infringement. Available at: https://apnews.com/article/disney-universal-midjourney-copyright-lawsuit-722b1b892192e7e1628f7ae5da8cc427, 11. 6. 2025.

Chmielewski, D. 2025. Disney, Universal sue image creator Midjourney for copyright infringement. Available at: https://www.reuters.com/business/media-telecom/disney-universal-sue-image-creator-midjourney-copyright-infringement-2025-06-11/, 11. 6. 2025.

Chow, A. R. 2025. How the Disney-Midjourney Lawsuit Could Reshape the Battle Over AI and Copyright. Available at: https://time.com/7293362/disney-universal-midjourney-lawsuit-ai/, 11. 6. 2025.

Culliton, B. M. 2024. The generative AI pirate? The intersection of copyrights and generative AI in literary works. Cybaris: *An Intellectual Property Law R*eview, 15(3), pp. 251-283. https://doi.org/10.4324/9781003260127-2

Cyphert, A. B. 2024. Generative AI, plagiarism, and copyright infringement in legal. *Minnesota Journal of Law, Science and Technolog*y, 25 (Symposium issue), pp. 49-66. https://doi.org/10.2139/ssrn.4938701

Dermawan, A. 2023. Text and data mining exceptions in the development of generative AI models: What the EU Member States could learn from the Japanese "nonenjoyment" purposes?. *The Journal of World Intellectual Property*, 27, pp. 44-68. https://doi.org/10.1111/jwip.12285

Dornis, T. 2025. The Training of Generative AI Is Not Text and Data Mining. *European Intellectual Property Revie*w, 47(2), pp. 65-78.

Duffy, C. 2025. Disney and Universal team up to sue AI photo generator Midjourney, claiming copyright infringement. Available at: https://edition.cnn.com/2025/06/11/tech/disney-universal-midjourney-ai-copyright-lawsuit , 11. 6. 2025.

Dusollier, S. *et al.* 2025. Copyright and Generative Al: Opinion. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 16(1), pp. 121-127.

Dutta, M. 2025. The Legal Landscape of Copyright in the Era of Artificial Intelligence. *Indian Journal of Law and Legal Research*, 7(2), pp. 7768-7782.

European Commission, Directorate General for Communication,.2022. *Study on copyright and new technologies: copyright data management and artificial intelligence*. Brussels: Publications Office of the European Union.

Garon, J. 2023. The Revolution will be Digitized: Generative AI, Synthetic Media, and the Medium of Disruption. *Ohio State Technology Law Jour*nal, 20(1), pp. 139-223.

Gendron, V. 2024. A New Frontier: The Music Industry's Struggle against Generative AI. *University of Miami Business Law Review*, 33(1), pp. 161-192.

Guadamuz, A. & Cabell, D. 2014. Data Mining in UK Higher Education Institutions: Law and Policy. *Queen Mary Journal of Intellectual Property*, 4(1), pp. 3-29. https://doi.org/10.4337/qmjip.2014.01.01

Henderson, P. Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A. & Liang, P. 2023. Foundation Models and Fair Use. *Journal of Machine Learning Research*, 24(1), pp. 1-76. https://doi.org/10.2139/ssrn.4404340

Ivanović, S. 2021. Autorskopravo i pretraživanje teksta i podataka. *Zbornik radova Pravnog fakulteta u Niš*u, 60(92), pp. 59-78. https://doi.org/10.5937/zrpfn0-34357

Knibbs, K. 2025. Disney and Universal Sue AI Company Midjourney for Copyright Infringement. Available at: https://www.wired.com/story/disney-universal-sue-midjourney/, 11. 6. 2025.

Kollár, P. 2022. Mind if I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union. *SSRN Electronic Journal*, pp. 1-44. https://doi.org/10.2139/ssrn.3960570

Kuker, H. 2024. Inspiration Versus Infringement: Why The Right to Use Copyrighted Images for Referential Purposes by AI Should Not Be Held To Higher or

Different Standards Currently in Effect. *SSRN Electronic Jo*urnal, pp. 1-34. https://doi.org/10.2139/ssrn.4727896

LaPorta, L. 2024. Authorless AI: navigating copyright challenges of generative. *Penn Undergraduate Law Journal*, 11(1), pp. 64-88.

Lin, P. K. 2024. Retrofitting Fair Use: Art & Generative AI after Warhol. *Santa Clara Law Review*, 64(2), pp. 467-500. https://doi.org/10.2139/ssrn.4566945

Lobling, L., Handschigl, C., Hofmann, K. & Schwedhelm, J. 2023. Navigating the Legal Landscape: Technical Implementation of Copyright Reservations for Text and Data Mining in the Era of Al Language Models. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 14(4), pp. 499-409.

Lu, F. 2024. AI-Generated Content: Legal Challenges & Potential Reforms. *Lecture Notes in Education Psychology and Public Me*dia, 66(1), pp. 174-181. https://doi.org/10.54254/2753-7048/66/2024MU0035

Malevé, N. 2024. Lost in Compression: Models of Authorship in Generative AI. *Media Theory*, 8(1), pp. 205-228. https://doi.org/10.70064/mt.v8i1.1074

Margoni, T. & Kretschmer, M. 2022. A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. *GRUR International*, 71(8), pp. 685–701. https://doi.org/10.1093/grurint/ikac054

Mešević, I. R. 2024. Umjetna inteligencija i mediji: The New York Times Company protiv kompanija Open AI i Microsoft. In: Popović, D.V. (ed.), *Veštačka inteligencija - izazovi u poslovnom pravu*, Beograd: Pravni fakultet Univerziteta u Beogradu, pp. 53-79. https://doi.org/10.51204/Internet_Dijalog_2403A

Moerland, A. 2024. Tekst- en datamining over de grens. Ars Aequi, *Special issue*, pp. 682-681.

Montgomery, B. 2025. Disney and Universal sue AI image creator Midjourney, alleging copyright infringement. Available at: https://www.theguardian.com/technology/2025/jun/11/disney-universal-ai-lawsuit, 11. 6. 2025.

Murray, M. D. 2023. Generative AI Art: Copyright Infringement and Fair Use. *SMU Science and Technology Law Rev*iew, 26(2), pp. 259-316. https://doi.org/10.25172/smustlr.26.2.4

Napitupulu, P. A., Sinaga, C. A. F. & Hasugian, A. L. P. 2023. The Implication of Generative Artificial Intelligence towards Intellectual Property Rights (Examining the Multifaceted Implications of Generative Artificial Intelligence on Intellectual Property Rights). *West Science Law and Human Rights*, 1(3), pp. 274-284. https://doi.org/10.58812/wslhr.v1i04.330

Orlando, S. & Casalini, L. 2023. The inexistence of an exclusive right of Text and Data Mining (TDM) as a matter of EU copyright law. *SSRN Electronic Journal*, pp. 1-34. https://doi.org/10.2139/ssrn.4578122

Perišić, J. Ć., 2025. Razvoj industrije veštačke inteligencije u Evropskoj uniji: izazovi i perspektive. In: Šarkić, N. (ed.), *Zbornik radova sa Savetovanja pravnika: [XXVIII Savetovanje Budvanski pravnički dani,* Budva, June 2025. Belgrade: Udruženje pravnika Srbije, pp. 613-628.

Popović, D. V. 2024. Ograničenja autorskog i srodnih prava: prepreka ili podsticaj razvoju veštačke inteligencije. In: Popović, D. V. (ed.), *Veštačka inteligencija - izazovi u*

poslovnom pravu, Beograd: Univerzitet u Beogradu Pravni fakultet, pp. 33-52. https://doi.org/10.51204/Internet_Dijalog_2402A

Radovanović, S. 2018. Otvorena nauka i autorsko pravo, s posebnim osvrtom na Predlog direktive EU o autorskom pravu na jedinstvenom digitalnom tržištu. In: Popović, D. V. (ed.), *Intelektualna svojina i internet*. Beograd: Pravni fakultet Univerziteta u Beogradu, pp. 21-36. https://doi.org/10.51204/Internet_Dijalog_1802A

Rosati, E. 2019. Copyright as an Obstacle or an Enabler: A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity. *Asia Pacific Law Review*, 27(2), pp. 198-217. https://doi.org/10.1080/10192557.2019.1705525

Sag, M. 2023. Copyright Safety for Generative AI. *Houston Law* Review, 61(2), pp. 295-348. https://doi.org/10.2139/ssrn.4438593

Sag, M. 2024. Fairness and Fair Use in Generative AI. *Fordham Law Review*, 92(5), pp. 1887-1922.

Senftleben, M. 2022. Compliance of National TDM Rules with International Copyright Law – An Overrated Nonissue?. *International Review of Intellectual Property and Competition Law*, 53(10), pp. 1477-1505. https://doi.org/10.1007/s40319-022-01266-8

Sobel, B. L. 2024. Elements of Style: Copyright, Similarity, and Generative AI. *Harvard Journal of Law & Technology*, 38(1), pp. 49-106. https://doi.org/10.2139/ssrn.4832872

Szkalej, K. 2025. The Paradox of Lawful Text and Data Mining? Some Experiences from the Research Sector and Where We (Should) Go from Here. *GRUR International*, 74(4), pp. 307-319. https://doi.org/10.1093/grurint/ikaf029

Toth, A. K. 2019. Algorithmic Copyright Enforcement and AI: Issues and Potential Solutions, through the Lens of Text and Data Mining. *Masaryk University Journal of Law and Technology*, 13(2), pp. 361-388. https://doi.org/10.5817/MUJLT2019-2-9

Widła, B. 2025. Thou Shalt Not Conduct Research on Software? Text and Data Mining of Computer Programs in the Current EU Copyright Framework. *GRUR International*, 74(1), pp. 3-18. https://doi.org/10.1093/grurint/ikae147

Živković, A. Đ., 2024. Pravni okvir zaštite računarskih programa sa posebnim osvrtom na veštačku inteligenciju „ChatGPT". *Strani pravni život*, 68(3), pp. 317-338. https://doi.org/10.56461/SPZ_24301KJ

Legal Sources

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. 2019. Official Journal L 130, 17 May 2019, pp. 92–125.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. 2024. Official Journal L, pp. 2024/1689, 12 July 2024.

The Code of Law of United States of America. 2024.

Case Law

Kneschke v. LAION, Hamburg Regional Court (Landsgericht Hamburg) of 27 September 2024, (Urteil 310 O.22723), available at https://openjur.de/u/2495651.html [Accessed 8 December 2025].

Getty Images vs. Stability AI, 2025, High Court of Justice Business and Property Courts of England and Wales of 4 November 2025, Case No: IL-2023-000007, EWHC 2863 (Ch), available at https://www.judiciary.uk/wp-content/uploads/2025/11/Getty-Images-v-Stability-AI.pdf [Accessed 8 December 2025].

GEMA vs OpenAI, Munich Regional Court (Landsgericht München), First Chamber, of 11 November 2025, (Urteil 42 O 14139/24), available at https://openjur.de/u/2537592.html [Accessed 8 December 2025].