Copy, Paste, and Generate: Copyright Law and Fair Use in the Age of Artificial Intelligence

Andy Benzo[8]

Independent Researcher

Recibido: 15/05/2025

Aceptado: 13/10/2025

✉ andybenzo@jurismentis.com

[8] ***Andy Benzo*** is a legal language expert and legal translator with over 35 years of experience specializing in English-Spanish translations. She holds degrees in Law and Translation and a Paralegal Degree from California State University, San Marcos. A certified ATA member since 1999, she now serves as President for the 2025-2027 term.

Abstract

This article examines whether the U.S. doctrine of fair use can adequately address the legal and ethical challenges posed by the training of generative artificial intelligence (GenAI) systems. The key research question driving this analysis is: *Can fair use, as currently interpreted, provide a coherent and equitable framework for addressing the massive, automated ingestion of copyrighted works for the AI model development?* This research is guided by two hypotheses. First, the transformative-use doctrine, while important in modern fair use analysis, remains insufficiently defined and inconsistent to address the functional and non-expressive nature of AI training. Second, that the current litigation landscape, as seen in *The New York Times v. Microsoft Corporation and OpenAI, et al.* (case no. 1:23-cv-11195 Southern District of New York) "*OpenAI case*" and *Dow Jones & Company, Inc. et al v. Perplexity AI, Inc.* (case no. 1:24-cv-07984 Southern District of New York) "*Perplexity case*", indicates an urgent need for legislative clarification to harmonize innovation incentives with copyright protection in the age of machine learning. The article uses a doctrinal legal analysis, examining case law, statutes, and policies, to show how fair use law has changed and how it applies to new AI technology. This methodological approach contextualizes the dispute within its historical origins and current policy ramifications, offering a cohesive legal and ethical framework for evaluating the limits of fair use in the age of generative AI.

Keywords*:* generative AI, copyright, authorship, fair use, AI training, transformative use, market substitution, AI litigation.

Resumen

Este artículo examina si la doctrina estadounidense del uso justo puede abordar adecuadamente los desafíos legales y éticos que plantea el entrenamiento de sistemas de inteligencia artificial generativa (GenAI). La pregunta clave de la investigación que impulsa este análisis es: *¿Puede el uso justo, tal como se interpreta actualmente, proporcionar un marco coherente y equitativo para abordar la ingesta masiva y automatizada de obras protegidas por derechos de autor para el desarrollo del modelo de IA?* Esta investigación se guía por dos hipótesis. En primer lugar, la doctrina del uso transformativo, aunque es importante en el análisis moderno del uso justo, sigue siendo insuficientemente definida e inconsistente para abordar la naturaleza funcional y no expresiva del entrenamiento de IA. En segundo lugar, que el panorama actual de los litigios, como se ve en *The New York Times v. Microsoft Corporation y OpenAI, et al.* (caso no. 1:23-cv-11195 Distrito Sur de Nueva York) "*Caso OpenAI*" y *Dow Jones & Company, Inc. et al v. Perplexity AI, Inc.* (caso No. 1:24-cv-07984 Distrito Sur de Nueva York) "*Caso Perplexity*", indica una necesidad urgente de aclaración legislativa para armonizar los incentivos a la innovación con la protección de los derechos de autor en la era del aprendizaje automático. El artículo utiliza un análisis legal doctrinal, examinando la jurisprudencia, los estatutos y las políticas, para mostrar cómo ha cambiado la ley de uso justo y cómo se aplica a la nueva tecnología de IA. Este enfoque metodológico contextualiza la disputa dentro de sus orígenes históricos y ramificaciones políticas actuales, ofreciendo un marco legal y ético cohesivo para evaluar los límites

del uso justo en la era de la IA generativa.[9]

Palabras clave: IA generativa, derechos de autor, autoría, *fair use*, entrenamiento de IA, uso transformativo, sustitución de mercado, litigios relacionados con IA.

[9] En muchos países de habla hispana, el concepto de uso justo no existe per se en la ley de derechos de autor. En cambio, se reconocen "límites y excepciones al derecho de autor" (Ley de Propiedad Intelectual de España o Ley 11.723 de Argentina). Por lo tanto, al traducir o interpretar el uso legítimo en contextos legales, es importante aclarar que se refiere específicamente a la doctrina de los EE. UU. bajo 17 U.S.C. § 107.

Introduction

By examining the consistency and shortcomings of applying fair use precedents to contemporary Generative AI litigation, this research argues that the legislative action is needed to establish clear boundaries for machine learning models.

This research uses doctrinal legal analysis as its primary methodological framework. Doctrinal research, grounded in the interpretation of statutes and judicial decisions, remains an essential component of legal scholarship. It is especially useful for exploring how the fair use doctrine is changing in relation to generative artificial intelligence. (GenAI). This approach analyzes the connection between statutory language, judicial precedent, and the policy objectives embedded in copyright law, aiming to identify both the consistencies and shortcomings that emerge when traditional legal principles are applied to modern technology contexts.

The methodological procedure consists of three steps. A thorough analysis of key judicial decisions reveals the foundation for current AI-related lawsuits: *Sony Corp. of America v. Universal City Studios, Inc. (1984*) established fair use for personal home recording and a safe harbor for technology with legal uses; *Campbell v. Acuff-Rose Music, Inc. (1994)* recognized commercial parodies as fair use if transformative; *Authors Guild, Inc. v. Google, Inc. (2015)* found Google's book-scanning project to be transformative fair use; and *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith (2023)* limited fair use when a work's commercial purpose outweighs its new meaning.

Second, the analysis incorporates comparative case evaluation, examining ongoing litigation such as *The New York Times v. Microsoft Corporation and OpenAI*,

as well as *Dow Jones & Company, Inc. v. Perplexity AI*. These cases serve as modern assessments of the applicability of fair use in relation to mass data ingestion and algorithmic transformation.

Finally, the study considers the broader impact of judicial reasoning on creative industries, journalism, and innovation policy. This multidisciplinary approach links law, technology, and ethics, fostering a nuanced understanding of how fair use operates at the intersection of human authorship and machine learning.

This tripartite framework, doctrinal, comparative, and policy-oriented, aims to clarify the doctrinal limits of fair use while providing a well-informed basis for legal and regulatory reform in the United States.

The Fair Use Doctrine

The fair use doctrine, currently caught in a "doctrinal crossroads" between the broad functional transformation principles established in cases like *Authors Guild v. Google* and the restrictive market-substitution focus of *Andy Warhol Foundation v. Goldsmith*, is structurally inadequate to govern the systemic and automated reproduction inherent in Generative AI, necessitating judicial reform that explicitly differentiates between non-expressive computational training and expressive, market-competitive content generation.

The fair use doctrine is very important in U.S. copyright law because it balances the exclusive rights of authors with the public's interest in the dissemination of information and the growth of creativity. Fair use is a broad limit on copyright exclusivity that is spelled out in Section 107 of the Copyright Act of 1976. It lets anyone use protected works without permission if it is in the public interest. Its open-

ended wording shows that Congress wanted to keep things flexible as technologies change. This has benefits, but it also makes things harder to understand when applied to new situations like generative artificial intelligence (GenAI).

The concept of fair use originates from equitable common-law principles delineated in early nineteenth-century jurisprudence, particularly in *Folsom v. Marsh* (1841), where Justice Story initiated the fundamental examination of "the nature and objects of the selections made." This adaptable, case-specific methodology subsequently transformed into a formalized, four-factor balancing test under 17 U.S.C. § 107.

The law requires courts to consider several factors when determining fair use: the intent and nature of the utilization, the type of work that is protected by copyright, the size and importance of the part used, and the impact of the utilization on the prospective market. This multifactor analysis is neither automated nor exhaustive; it permits judicial interpretation in context, enabling courts to adapt the doctrine to accommodate emerging technologies and media. This adaptability has resulted in "doctrinal indeterminacy," where courts disagree on factors like transformative purpose and market impact.

Throughout the 20th century, numerous instances arose where copyright law conflicted with emerging technologies. Every instance of photocopying, home recording, digital sampling, or web searching was evaluated for its compliance with fair use. In the landmark case ***Sony Corp. of America v. Universal City Studios, Inc.* (1984),** it was ruled that recording TV shows for later viewing ("time-shifting") was a fair use. This case also introduced the idea of technological neutrality: if a device has a lot of non-infringing uses, its distribution should not be considered infringing.

In ***Campbell v. Acuff-Rose Music, Inc***. (1994), ten years later, the first principle was changed to include the idea of transformative use, which is a use that adds new meaning, expression, or message to the original work. The Court rejected the idea that commercial use always negates fair use, instead highlighting the importance of creative transformation. This line of thinking made large-scale copying legal when it was done for analytical or functional purposes. AI engineers currently use this idea to explain why they need to consume data for AI training.

The case of ***Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith (2023)*** demonstrated that the law was getting stricter. The Supreme Court clarified the transformative test by emphasizing the commercial purpose of the secondary use and its similarity to the original in the market. The Court stated that granting a magazine authorization to utilize a Warhol print was equivalent to permitting the use of a Goldsmith photograph, which contravened fair use principles. This shift toward market substitution makes it harder to defend GenAI training that only relies on functional transformation.

The Modern Crossroads

The combination of *Authors Guild* and *Warhol* produces a doctrinal crossroads. The first group supports innovation by saying that functional transformation is fair, whereas the second group preserves authors' markets by limiting how broadly transformation can be read. GenAI systems converge at this point: they analyze language functionally but produce expressive content that can compete with original works.

This tension illustrates the adaptability of fair use doctrine. The concept,

designed for isolated instances of copying, struggles to govern systemic, automated reproduction. The case-by-case framework lacks explicit guidance for evaluating both analytical and generative usage. The article argues that, while historically adaptable, the fair use doctrine is doctrinally insufficient to address the epistemic and economic disruptions caused by GenAI.

Acknowledging these limitations calls for a reevaluation of fair use from both practical and ethical perspectives. Courts must differentiate between non-expressive computational uses, such as data analysis and pattern recognition, and expressive market applications, such as content generation. In a normative sense, they must reaffirm the constitutional goal of copyright, "to promote the progress of science and useful arts", by ensuring that technological innovation doesn't weaken the incentives for people to be creative.

The following sections explain this idea by examining how ongoing lawsuits between major media companies and AI developers challenge the fairness of fair use, highlighting the urgent need for legal and judicial reform.

Analysis

The high-stakes lawsuits initiated by *The New York Times* and *Dow Jones* demonstrate that Generative AI's dual challenge of unauthorized data ingestion and market-competitive expressive output requires the U.S. judiciary to reform the fair use doctrine by establishing a clear, context-sensitive distinction between protected analytical transformation and unprotected commercial substitution.

The conflict between traditional media companies and tech innovators has tested the fair use doctrine like never before. The *New York Times Company v.*

Microsoft Corporation and OpenAI, and *Dow Jones & Company, Inc. v. Perplexity AI, Inc.*, are the first major cases to examine the legality of using large amounts of data to train AI under U.S. copyright law. These cases illuminate the broader legal and ethical question of whether GenAI's use of copyrighted material is transformative or constitutes a systemic infringement.

The New York Times vs Microsoft Corporation and OpenAI

The *New York Times (NYT*) lawsuit, filed in the Southern District of New York on December 27, 2023, was a turning point in copyright cases involving generative AI. The complaint says that OpenAI and Microsoft "copied and used millions of The Times's copyrighted works without permission" to train big language models like the GPT series. The NYT states that this use is deliberate rather than accidental, noting that its journalism was "given particular emphasis" in the training datasets, which indicates that it recognized and relied on its expressive value.

The lawsuit states that the defendant has violated the law in two ways. First, at the input stage, it says that duplicating full articles to train a model is a direct infringement. Second, at the output stage, it states that AI models "memorize" and recreate expressive information, often word-for-word, which produces derivative works that compete directly with the NYT's own works. The complaint document presents instances in which ChatGPT and Microsoft's Copilot reproduce *New York Times* articles with near-identical content, indicating that users may not require access to the original pieces.

The defendants' response cites *Authors Guild v. Google*, asserting that AI training constitutes a transformative analytical use equivalent to creating a search index.

The objective is said to be functional, to understand the mechanics of language and the interconnectedness of words, rather than expressive. This viewpoint stresses that AI training does not convey or mimic human-generated expression but enables computational understanding of language.

Judge Sidney H. Stein's initial rulings in early 2025 did not dismiss the primary copyright allegations. This indicated that the court was cautious regarding an expansive interpretation of transformative usage. The case has been consolidated with analogous claims from other publishers into a multidistrict litigation titled: In re: *OpenAI Copyright Infringement Litigation* (1:25-md-03134, S.D.N.Y.). The discovery phase has proven notably contentious, particularly over the retention of chat logs and the admissibility of evidence pertaining to retained material.

The New York Times case illustrates the contentious aspects of contemporary fair use analysis. The defendants argue that technology is necessary and changes how things work; the plaintiffs argue that it merely copies how things are expressed and takes over the market. The outcome will depend on how the court interprets the relevant "use." If "use" is limited to the analytical training process, the defense could prevail under *Authors Guild* rationale; on the other hand, if it encompasses both training and expressive outputs, the market-centric reasoning of *Warhol v. Goldsmith* will apply, favoring the plaintiffs.

Dow Jones & Company, Inc. v. Perplexity AI, Inc.

In June 2024, Dow Jones, the parent company of *The Wall Street Journal* and the *New York Post*, initiated legal action against Perplexity AI in the same jurisdiction. The complaint alleged that Perplexity's "answer engine," a retrieval-augmented

generation (RAG) technology, routinely retrieved and reproduced copyrighted articles without authorization. RAG systems differ from conventional LLMs in that they acquire and integrate real-time external data to generate summaries and snippets that often include verbatim content from news sources.

Dow Jones's claim centers on direct market substitution. Perplexity's marketing slogan, "Skip the Links," encourages users to utilize AI-generated responses rather than visiting the source publisher websites. This diverts traffic from those websites and the associated advertising revenue. The case additionally asserts that Perplexity damaged Dow Jones's reputation by occasionally fabricating "hallucinated" or erroneous assertions attributed to Dow Jones publications. This undermined the brand's credibility.

The case is currently before Judge Katherine Polk Failla. The plaintiffs want statutory damages, an injunction, and, most critically, the removal of Perplexity's RAG databases containing pirated materials. The defendants, in their motion to dismiss, reference *Authors Guild*, asserting that their use of news items for "summarization and contextual retrieval" serves a transformative informational purpose.

However, this defense is weakened by the explicit evidence of market substitution. Courts have long treated the fourth factor, the effect on the potential market, as "the single most important element of fair use." Here, Perplexity's business model appears to replace rather than complement the original market. Furthermore, its practice of showing large passages goes beyond the limited, non-substitutive use that judges allowed in *Authors Guild*.

This study argues that the *Perplexity* case highlights an important doctrinal difference. While training a model may be seen as a transformative analytical act, using

that model in a way that competes with copyrighted expression blurs the line between making tools and replacing content. The defendants' reliance on technological transformation cannot overcome the clear commercial equivalence between their outputs and the plaintiffs' works.

These two lawsuits illustrate the dual aspects of GenAI's legal challenges: the input issue (the permissibility of copying for training purposes) and the output issue (the impact of generated content as a replacement). They illustrate how courts must navigate two contradictory precedents —*Authors Guild* and *Warhol* —that pull in opposite directions.

From a policy perspective, these instances represent a pivotal moment in achieving the right balance between safeguarding and innovation. If courts favor AI developers, they could hurt the economic foundations of journalism and creative fields. If they only support rights holders, they might slow technological innovation and make it harder for people to access data.

This author believes that the answer is not to be strict about doctrine, but to have a new, context-sensitive definition of fair use that makes a clear distinction between analytical transformation and expressive substitution.

Discussion

The application of the four statutory fair use factors (U.S.C.) to Generative AI (GenAI) creates an unworkable conceptual tension, as the allowance for functional innovation (Factors 1 and 3) is directly antagonized by the protection of expressive rights and market integrity (Factors 2 and 4), rendering the analog-era doctrine structurally insufficient to govern systemic, automated reproduction.

The conceptual tension in fair use jurisprudence is particularly pronounced when the four statutory elements of 17 U.S.C. §107 are applied to the context of training and operation of generative artificial intelligence (GenAI) systems. The factors—purpose and character of the usage, nature of the copyrighted work, amount and substantiality of the portion used, and impact on the potential market—were intended for distinct, human-centric acts of reproduction. This presents profound analytical and philosophical challenges about their potential application in algorithmic, large-scale data input and synthetic output generation.

This section examines each factor from two perspectives: doctrinal interpretation and technological reality. It shows how recent court decisions have both clarified and narrowed the potential scope of fair use in AI training and deployment.

The first factor is the "purpose and character" of the use, and it has historically developed as the most contentious and conceptually dynamic. The Supreme Court's definition of transformative use in *Campbell v. Acuff-Rose Music, Inc*. (1994) shifted the emphasis from mechanical substitution to creative reinterpretation. According to this standard, a secondary use is transformative if it changes the original work "with new expression, meaning, or message."

AI engineers frequently reference this precedent to assert that training a model constitutes a transformative process. They contend that the utilization of huge textual datasets serves a functional purpose rather than an expressive one: the model does not reproduce meaning but statistically extracts patterns of syntax, semantics, and style to enable the synthesis of novel material. This perspective suggests that instructing an AI parallels Google's creation of a searchable corpus in *Authors Guild v. Google* (2015).

This was termed "highly transformative" as it created an informational tool rather than a market substitute.

This approach, however, faces significant challenges as compared to *Warhol Foundation v. Goldsmith* (2023). The Court in *Warhol* limited the transformative analysis by asserting that even if a secondary work has a new message or look, it fails the first criterion if it serves a "substantially similar commercial purpose" to the original. This new way of looking at things shifts the focus from the vague idea of "new meaning" to the commercial function of use.

Applying this approach to GenAI, training may appear to be a significant alteration; nonetheless, the outputs generally replicate the functions of the original works on which the models were trained, such as composing essays, articles, or artistic creations. From this perspective, the distinction between transformative tool creation and competitive replication disintegrates. The main issue now relies more on commercial substitution than technological innovation, making AI developers' reliance on *Authors Guild* less stable after *Warhol.*

The second factor examines the nature of the work being replicated. Judicial systems afford greater protection to highly creative and expressive works than to those consisting just of facts or information. This distinction is problematic in the context of AI training, as most datasets consist of hybrid materials—journalistic articles, academic essays, and creative works that combine factual reporting with interpretative expression.

Developers assert that AI models primarily identify linguistic patterns, which are unprotectable elements such as thoughts or facts. Rights holders contend that the arrangement of words and their phonetic qualities are essential components of artistic

expression. In its May 2025 report, the Copyright Office rejected the argument that training AI models is simply a statistical process. The office stated that language models learn "the essence of linguistic expression," including how words are selected and arranged, rather than just processing words. Consequently, this factor is more likely to safeguard works that are expressive, such as literature, commentary, and artistic journalism.

The third factor, "the amount and substantiality of the portion used," considers both the amount and the quality of the portion used. In the past, courts have tolerated limited copying when necessary for a transformative purpose. However, as stated in *Authors Guild*, complete reproduction may qualify as fair use if it serves a transformative purpose.

AI developers use this example to argue that processing entire corpora is essential for effective AI training and understanding linguistic context. This defense is doctrinally plausible, but it raises additional questions about scale and need. GenAI models can replicate and alter expressive material on a large scale, unlike Google's digitization effort, which only displayed small "snippets." The outcome is not just incidental exposure but also the potential internalization of creative expression within the model's parameters.

This author observes that the third factor now exposes a structural conflict between the qualitative norm of "heart of the work" and the quantitative reality of full replication inherent in machine learning. Courts must ascertain whether such comprehensive intake, however functionally essential, disrupts the equilibrium that fair use aims to uphold.

The fourth factor examines whether the secondary use reduces or dilutes the original's potential market. The current GenAI litigation has elevated this issue significantly across the legal and business sectors.

The plaintiffs in *The New York Times* and *Perplexity* claims assert that they have incurred financial losses and that the market structure is deteriorating due to AI platforms diverting audience attention, advertising revenues, and subscription value from news websites without compensation. This market impact goes beyond mere substitution; it also includes the elimination of opportunities for licensing AI training, a sector that is already expanding through partnerships between publishers and technology firms.

AI enterprises assert that allegations regarding diminished licensing markets are circular: if the usage is equitable, a license is unnecessary. However, courts are beginning to question this perspective. In *Kadrey v. Meta (2025)*, Judge Vince Chhabria suggested that the substantial capacity of generative AI (GenAI) to saturate the market with paraphrased or derivative content could constitute "market dilution," a new form of injury under the fourth fair use factor.

This suggests that the market-effect element strongly argues against fair use in how AI is trained today. The combination of direct substitute, lost licensing revenue, and damage to reputation creates an overall economic impact that contradicts the fair use principle of fairness.

The four factors collectively demonstrate the fragility of the fair use concept in the era of artificial intelligence. Factors 1 and 3 facilitate functional innovation, whereas Factors 2 and 4 safeguard expression and market integrity. The resultant balance

increasingly favors restraint over expansion.

This author believes that the fair use doctrine, initially formulated for analog contexts, fails to meet the interpretive requirements imposed by systemic, automated, and probabilistic reproduction. In the absence of legal modifications, courts will continue to render illogical rulings, oscillating between technical exceptionalism and market conservatism.

Implications

The legal battle over the fair use doctrine in Generative AI (GenAI) goes beyond simple legal interpretation, embodying a key ethical and societal challenge that, by enabling the widespread economic displacement of human creators and threatening the trustworthiness of democratic institutions like journalism, requires a proactive legal and ethical approach to ensure that technological progress respects human creativity. The ongoing conflict between copyright law and generative artificial intelligence (GenAI) extends beyond mere legal interpretation. It addresses the core concept of intellectual property: how society delineates authorship, creativity, and equity in an era where robots can replicate and generate novel creations. The legal discourse around fair use has substantial ethical, economic, and democratic implications.

The primary objective of copyright law is to achieve a balance between fostering creativity and ensuring the accessibility of knowledge to all individuals. GenAI disrupts this equilibrium by rendering the size and value extraction imbalanced. The media and publishing industries invest substantial time, resources, and personnel in creating original content, but AI businesses generate revenue by utilizing those works without authorization to produce their own products.

This conflict is not merely a concept; it manifests in actual economic displacement. The lawsuits filed by *The New York Times* and *Dow Jones* underscore a structural market distortion: AI tools that summarize or replicate journalistic content diminish website traffic, impair advertising revenue, and undermine the subscription-based model that sustains quality journalism.

Ultimately, this disparity may engender a detrimental feedback loop: diminished funding results in fewer distinctive creations, hence degrading the quality of the data upon which AI systems depend. The issue extends beyond financial loss; it also pertains to ensuring that human innovation remains a public asset.

The ethical dimension of the fair use controversy revolves around authorship, intent, and accountability. In 2023, the U.S. Copyright Office reaffirmed that copyright protection is limited to works of human authorship. This principle establishes a clear distinction between human creativity, defined by intent, context, and moral ownership, and machine-generated output, which lacks these characteristics.

However, GenAI technologies obscure this distinction. The distinction between inspiration and appropriation becomes ambiguous when outputs closely resemble individual expressions to the extent that they may serve as substitutes. AI-generated texts may appear original; yet, they are really compilations of human-authored content that have been statistically amalgamated without context or consent. This technique raises issues with moral responsibility: if robots learn from human input, what obligations do their creators bear?

From an ethical standpoint, the concept of fair use—an equitable principle of rationale—was never designed to permit broad copying of human expression on a grand

scale. The ethical foundation suggests reciprocity and proportionality, both of which are undermined when the "user" is a non-human entity representing a corporation. This suggests that the transformation of fair use from a tool for equity to a protection for corporate automation undermines public trust in the fundamental fairness of copyright law.

GenAI influences society in ways that extend beyond mere politics and public discourse. Journalism, sometimes referred to as the "fourth estate," must maintain financial independence and public trust to fulfill its role as an effective watchdog. AI systems that assimilate and reformat news articles can simultaneously cause harm to both entities involved.

Unauthorized reproduction of news articles reduces revenue and makes it harder to identify the original source, thereby undermining openness. In an era characterized by pervasive misinformation, it is crucial to ascertain the origins of information. AI systems that paraphrase or fabricate news articles risk obscuring the distinction between authentic journalism and fictitious narratives. This problem, seen in the Perplexity AI instance, suggests that unchecked generative outputs could jeopardize the epistemic structure of democratic societies.

The ethical inquiry pertains not only to ownership but also to the integrity of the information. The fair use doctrine that permits extensive decontextualization of journalism inadvertently undermines its constitutional role in promoting an informed public.

AI developers emphasize that restricting access to training data would hinder research and impede the dissemination of technology to everybody. They reference the

case of *Sony v. Universal City Studios* (1984) to argue that copyright should not impede technological progress. From this viewpoint, GenAI represents a logical advancement in humanity's persistent endeavor to automate knowledge.

This argument does have some merit. Historically, open access to data has facilitated scientific advancement and innovative synthesis. GenAI has the power to improve things by advancing research, education, and accessibility. But innovation without responsibility can lead to exploitation. The challenge, then, is to tell the difference between innovation that adds to our shared knowledge and appropriation that devalues our culture.

To achieve long-term balance, the interests of creators, consumers, and innovators must align. Copyright law has to shift from a reactive to a proactive system that can keep up with technological changes. This means that both the law and the way businesses operate should incorporate ethical values such as consent, openness, and proportionality.

As the evidence indicates, the current AI litigation teaches us something more important: copyright is not meant to stop technology, but to make it more human. Legal and ethical reform should not seek to impede AI progress but rather to guarantee that innovation respects the creative and moral integrity of human authorship.

Conclusion

The convergence of generative artificial intelligence (GenAI) with copyright law illustrates the flexibility and vulnerability of the fair use theory. This essay aimed to address a fundamental research inquiry: Is the law of fair use sufficiently equipped to address the issues presented by generative AI training?

An analysis of doctrinal, technological, and policy dimensions leads to a complex conclusion: the current fair use framework insufficiently accommodates the systemic realities of AI training and deployment. Although it remains a crucial safeguard for creativity and innovation, its principles, particularly the transformative-use criterion, inadequately address automated, large-scale copying motivated by commercial interests.

This study was guided by two central hypotheses: first, that the transformative-use theory does not offer sufficient clarity to effectively address the complexities of AI training; and second, that the current landscape of litigation demonstrates a clear need for greater legal precision and guidance in this area.

Both possibilities have been confirmed. The ruling in *Authors Guild v. Google* (2015) provided a functionalist rationale for fair use, validating extensive digitalization for non-expressive objectives. However, *Warhol Foundation v. Goldsmith* (2023) shifted the case's emphasis to market replacement, illustrating the economic fragility of the creative industries. The instances of *The New York Times v. Microsoft and OpenAI* and *Dow Jones v. Perplexity AI* demonstrate that the discord between these two concepts has reached an irreconcilable stage.

The primary strength of the fair use doctrine lies in its flexibility; nonetheless, the same characteristic has emerged as its most significant weakness in the era of machine learning. When using instruments designed for human, case-specific disputes, courts must navigate unfamiliar territory. If the laws remain unamended, the outcomes will persist in dysfunction, resulting in ambiguity for both creators and innovators.

This author asserts that the path ahead involves redefining fair use as a balanced

mechanism for holding technology accountable, one that honors human ingenuity while promoting responsible innovation. The subsequent recommendations delineate certain measures to enhance the U.S. legal and policy framework in this context.

Recommendations

Congress should consider **updating Section 107 of the Copyright Act** to explicitly address the use of copyrighted materials in machine learning and generative AI training. The statute's current open-ended language invites inconsistent judicial interpretation, particularly regarding the distinction between analytical and expressive uses.

Specific reforms might include defining "computational use" as a distinct category of fair use, which would be limited to analytical or non-expressive purposes and established with clear boundaries on data retention and derivative outputs. Additionally, these reforms could codify disclosure requirements, obligating AI developers to document the sources of their datasets and, where appropriate, obtain licenses for datasets that contain expressive works. To further enhance the framework, it is important to clarify the analysis of market harm by formally recognizing licensing markets for AI training as legitimate and protectable under Factor 4 of Section 107. These amendments would provide clarity and reduce the doctrinal uncertainty that currently burdens both courts and stakeholders.

Although updating legislation is essential, judicial interpretation continues to play a vital role in determining how fair use is applied in practice. Courts should employ a more systematic approach when evaluating cases related to AI, which involves clearly differentiating between the analysis of data used for AI training (input) and the uses of AI-generated content (output). This structured framework would also incorporate

contextual proportionality in assessing the third fair use factor, ensuring that the amount of copied material is truly necessary for the intended purpose and that adequate measures are in place to prevent unauthorized reproduction of expressive works. Additionally, the transformative-use analysis should be broadened to include not just the new purpose served by the AI, but also the extent of transparency and control exercised by the developer or user, further safeguarding creative integrity in the evolving digital landscape. Such guidelines would enable courts to balance innovation incentives with market protection, maintaining consistency across jurisdictions.

Transparency is the cornerstone of ethical AI development, yet the opacity of AI training processes—often referred to as the "black box" problem—undermines both legal enforcement and public trust. To address these concerns, policymakers should mandate the disclosure of training sources, work categories, and provenance metadata to ensure dataset transparency. Additionally, implementing audit mechanisms for independent verification of compliance with copyright and privacy standards, as well as establishing traceability systems that enable users and rights holders to determine whether specific content was derived from protected materials, are essential steps. These measures would not only facilitate fair use analysis but also strengthen AI governance by aligning it with broader principles of accountability and due process.

Legal reform must be accompanied by cultural and ethical change. Universities, corporations, and professional associations should incorporate AI ethics and copyright literacy into their education and training programs, ensuring that translators, writers, and journalists have the knowledge to understand how their works are used in AI contexts and are empowered to advocate for equitable treatment. Ethically, AI

developers should adopt the principle of "technological stewardship," taking responsibility to ensure that innovation serves the collective good. This involves proactive collaboration with creative communities and respect for the inherent value of human expression. As this author contends, the future of copyright is not solely a legal matter but also a moral one, raising the question of whether society values the act of creation as a fundamentally human endeavor or reduces it to a mere data point within an algorithmic system.

Reflection

The doctrine of fair use has always been an engine of progress, but its continued vitality depends on its capacity to evolve. Generative AI challenges the doctrine not because it is a new technology, but because it exposes the limits of legal concepts rooted in human intention. The law must therefore rise to meet the ethical scale of the technological revolution it confronts.

In supporting the hypotheses of this study, this article affirms that fair use, in its current doctrinal form, is not obsolete but insufficiently adapted. Legislative clarification, judicial discipline, and ethical foresight are essential to restore equilibrium between human creativity and machine innovation. Only by redefining fair use as a framework of responsible transformation can the United States continue to promote progress in the age of artificial intelligence.

References

AIAAIC. (n.d.). NYT orders Perplexity to stop misusing its content. https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/nyt-orders-perplexity-to-stop-misusing-its-content

Authors Alliance. (2024, February 29). Why fair use supports non-expressive uses. https://www.authorsalliance.org/2024/02/29/why-fair-use-supports-non-expressive-uses/

Authors Guild. (n.d.). Artificial intelligence. https://authorsguild.org/advocacy/artificial-intelligence/

Authors Guild, Inc. v. Google Inc., No. 13-4829-cv (2d Cir. Oct. 16, 2015). (n.d.). U.S. Copyright Office fair use summaries. U.S. Copyright Office. https://www.copyright.gov/fair-use/summaries/authorsguild-google-2dcir2015.pdf

California State University Stanislaus. (n.d.). Fair use 101: The four factors. https://www.csustan.edu/office-academic-technology/fair-use-101-four-factors

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994). (n.d.). U.S. Copyright Office fair use summaries. U.S. Copyright Office. https://www.copyright.gov/fair-use/summaries/campbell-acuff-1994.pdf

Center for Art Law. (2025, July). Unpacking the US Copyright Office's third report on generative AI. https://itsartlaw.org/2025/07/08/unpacking-the-us-copyright-offices-third-report-on-generative-ai/

Columbia University Libraries. (n.d.). Fair

use. https://copyright.columbia.edu/basics/fair-use.html

Congress.gov. (n.d.). Generative artificial intelligence and copyright law (LSB10922). https://www.congress.gov/crs-product/LSB10922

Cornell Law School Legal Information Institute. (n.d.). 17 U.S. Code § 107 – Limitations on exclusive rights: Fair use. https://www.law.cornell.edu/uscode/text/17/107

Ethical Web Data. (n.d.). Is web scraping legal? Navigating terms of service and best practices. https://ethicalwebdata.com/is-web-scraping-legal-navigating-terms-of-service-and-best-practices/

FindLaw. (n.d.). 17 U.S.C. § 107 – Limitations on exclusive rights: Fair use. https://codes.findlaw.com/us/title-17-copyrights/17-usc-sect-107/

FindLaw. (n.d.). Bell v. Wilmott storage services llc (2021). https://caselaw.findlaw.com/court/us-9th-circuit/2144129.html

Free.Law Court Listener (n.d.). Dow Jones & Company, Inc. et al. v. Perplexity AI, Inc., No. 1:2024cv07984. https://www.courtlistener.com/docket/69280523/dow-jones-company-inc-v-perplexity-ai-inc/

H & H Law. (2024, July). New York Times v. Microsoft [Complaint]. https://hh-law.com/wp-content/uploads/2024/07/New-York-Times-complaint.pdf

Harvard Law Review. (n.d.). Andy Warhol Foundation for Visual Arts, Inc. v. Goldsmith. https://harvardlawreview.org/print/vol-137/andy-warhol-foundation-for-visual-arts-inc-v-goldsmith/

IP Solved. (n.d.). AI & IP: Addressing the challenges of data scraping and the OECD view. https://ipsolved.com/news/posts/ai-ip-addressing-the-challenges-of-data-

scraping-and-the-oecd-view/

JD Supra. (n.d.). Fair use and AI training: Two recent decisions highlight the …. https://www.jdsupra.com/legalnews/fair-use-and-ai-training-two-recent-7198312/

LawInc. (2024, October). United States District Court Southern District of New York: Dow Jones & Company, Inc. et al. v. Perplexity AI, Inc. [Complaint]. https://www.lawinc.com/wp-content/uploads/2024/10/Perplexity-Lawsuit.pdf

Lutzker LLC. (n.d.). The *de minimis* doctrine: When is copying too trivial?. https://www.lutzker.com/the-de-minimis-doctrine-when-is-copying-too-trivial/

New York State Bar Association. (n.d.). Scratching the surface of IP rights: Data scraping and artificial intelligence. https://nysba.org/scratching-the-surface-of-ip-rights-data-scraping-and-artificial-intelligence/

Nolo. (n.d.). Fair use: The four factors courts consider in a copyright infringement case. https://www.nolo.com/legal-encyclopedia/fair-use-the-four-factors.htm

OpenAI. (n.d.). How we're responding to The New York Times' data demands in order to protect user privacy. https://openai.com/index/response-to-nyt-data-demands/

Oyez. (n.d.). Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith. https://www.oyez.org/cases/2022/21-869

Oyez. (n.d.). Campbell v. Acuff-Rose Music, Inc. https://www.oyez.org/cases/1993/92-1292

Oyez. (n.d.). Sony Corporation of America v. Universal City Studios,

Inc.. https://www.oyez.org/cases/1982/81-1687

Project MUSE. (n.d.). Generative AI and open access publishing: A new economic paradigm. https://muse.jhu.edu/article/961190

R&D World. (n.d.). Publisher Wiley tells AI developers: Get a license before training on our content. https://www.rdworldonline.com/publisher-wiley-tells-ai-developers-get-a-license-before-training-on-our-content/

Skadden, Arps, Slate, Meagher & Flom LLP. (2025, May). Copyright Office weighs in on AI training and fair use. https://www.skadden.com/insights/publications/2025/05/copyright-office-report

Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984). (n.d.). U.S. Copyright Office fair use summaries. U.S. Copyright Office. https://www.copyright.gov/fair-use/summaries/sonycorp-universal-1984.pdf

The Economic Times. (n.d.). Wall Street Journal sues Perplexity AI for copyright infringement. https://m.economictimes.com/tech/artificial-intelligence/wall-street-journal-sues-perplexity-ai-for-copyright-infringement/articleshow/114444319.cms

University of North Carolina at Charlotte. (n.d.). Fair use. https://library.charlotte.edu/research-write/copyright/fair-use

U.S. Copyright Office. (n.d.). Fair use index. https://www.copyright.gov/fair-use/

Washington and Lee University School of Law. (n.d.). Artificial Intelligence and Transformative Use After Warhol. *Washington and Lee Law Review*

Online, *81*(1). https://scholarlycommons.law.wlu.edu/wlulr-online/vol81/iss1/2/