EVALUATING THE APPLICATION OF FAIR USE AND FAIR DEALING PRINCIPLES IN THE CONTEXT OF TRAINING GENERATIVE ARTIFICIAL INTELLIGENCE (AI) MODELS

Shwetha Sunil Nambiar, LLM (IPR), Maharashtra National Law University, Mumbai

ABSTRACT

The training of Artificial Intelligence (hereinafter AI) models with copyrighted works has emerged as the most significant copyright concern plaguing courts, policymakers, and stakeholders worldwide.[1] While extensive literature addresses fair use and AI training in the United States Context, there exists a gap in scrutinizing how India's fair dealing framework under section 52 of the Copyright Act, 1957 applies to training AI models.[2] Recent opposing United States court decisions have further complicated the legal landscape with *Bratz v. Anthropic* (2025) and *Kadrey v. Meta* (2025) finding training of AI Models using copyrighted materials to be fair use[3] while *Thomson Reuters v. ROSS Intelligence* (2025) rejected the defense.[4] This paper addresses these gaps by conducting a systematic analysis of fair use and fair dealing doctrines operating in the US and Indian scenarios respectively applied to AI training[5]. This research critically evaluates whether the specific conditions enumerated in Section 52 of the Copyright Act, 1957 can adequately accommodate the nuances of training AI models in this context.[6] It analyzes patters and trends in the recent US case laws to ascertain factors explaining the diverse judicial decisions and using those as a foundation to propose legislative reforms in India that take up a balanced approach keeping in mind innovation, creator protection. This research also considers the present case of *ANI v. OpenAI* litigation pending in Delhi High

[1] Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOUS. L. REV. 101, 103 (2025); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. (forthcoming 2024); U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE: PART III – GENERATIVE AI TRAINING 3 (2025)

[2] Kailash Chauhan, *Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime*, 30(1) J. INTELL. PROP. RTS. 77, 78–79 (2025); Sapna Singh, *Feeding the Algorithm: Legal Challenges in AI Training Data for Educational Progress* 1–2 (2024).

[3] Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025); Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417 (N.D. Cal. June 25, 2025).

[4] Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc., No. 1:20-cv-00613 (D. Del. Feb. 11, 2025).

[5] Serena Lightstone, *Train or Restrain? Using International Perspectives to Inform the American Fair Use Analysis*, 44(3) NW. J. INT'L L. & BUS. 541, 560–75 (2024).

[6] Chauhan, *supra* note 2, at 79-82; Singh *supra* note, at 8-10.

Court, wherein India's application of fair dealing principle on AI Training will be tested by judiciary effectively for the first time.[7]

Introduction

The launch of Generative AI, like ChatGPT, Claude, Gemini has presented a global revolution in AI, colloquially, the 'AI Boom'. This rapid progression of use and reliance on AI has had an impact across sectors and affects human creativity and intellectual practice.[8] AI has improved over the past years rapidly where now generative AI can pass examinations[9], creating artwork that matches human made art.[10] AI has reshaped the economy in huge waves, the projected market growth is estimated to reach USD 180 billion by 2030;[11] Major AI companies have now become pivotal economic, social actors and their products have reformed employment, education, creative as well as legal fields.[12]

Understanding the legal implications that come up the growth in AI sector requires a baseline knowledge of how the system are developed and function. Training generative AI involves sequential phases- data collection, preprocessing, model training and output generation.[13] The training of AI requires developers to compile massive datasets and databases from open internet sources that contain datapoints such as books, articles, images, artwork, musical notes and compositions, software code and a myriad of copyrighted materials.[14] The data collected is the 'tokenized' (broken into smaller segments) and stored on servers where it is preprocessed and converted into numerical vectors.[15] When the AI is being trained, Machine Learning (ML) algorithms process these vectors, identify patterns and distributions within the data, encode the information into model's parameters that govern how inputs are transformed into outputs.[16] Scholarly debate persists whether this training process construes as 'consumption' of copyrighted materials, thereafter they disappear ad only a generalized pattern remains or whether it is 'memorized' wherein the AI model will later reproduce substantial portion of

[7] ANI Media Private Ltd. v. OpenAI Inc., CS(COMM) 1028/2024 (Delhi High Court, filed Dec. 2024) (pending); Singh, *supra* note 2, at 10–12
[8] Sapna Singh, *Feeding the Algorithm: Legal Challenges in AI Training Data for Educational Progress* 1 (2024).
[9] Daniel Martin Katz et al., *GPT-4 Passes the Bar Exam* (2023), https://ssrn.com/abstract=4389233; Yuntao Bai et al., *GPT-4 Technical Report*, arXiv preprint (2023).
[10] Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. (forthcoming 2024).
[11] GRAND VIEW RESEARCH, GENERATIVE AI MARKET SIZE, SHARE & TRENDS ANALYSIS REPORT (2023).
[12] Singh, *supra* note 1, at 1; Sag, *supra* note 3, at 2.
[13] Weiyi Xia, *Fair Use of Training Data in Generative Artificial Intelligence* 2–3 (2024).
[14] Sag, *supra* note 3, at 6.
[15] Jacqueline C. Charlesworth, *Generative AI's Illusory Case for Fair Use* 7 (2024).
[16] Xia, *supra* note 6, at 3; Sag, *supra* note 3, at 7.

training data.[17] This backend process is the foundation of generative AI, and when users provide prompts, the model will generate content and output by probabilistically predicting likely sequences taking into account learned data patterns.[18]

The legal implications of this training phase create a fundamental challenge. The initial training phase of copying, storing and processing of copyrighted works raises copyright concerns, such as right of reproduction under section 106(1) of the U.S. Copyright Act and similar provisions across the globe.[19] On the other hand, the AI companies assert that their conduct and process constitute fair use or fair dealing, since the training serves a different purpose, converse to enjoying creative expression availed in copyrighted works. Thus, not constituting as infringement of copyrights of the materials used in AI training.[20] The legal debate now surrounds how this claim made by AI companies have a direct conflict with established copyright doctrines that have historically protected authors rights and economic interest when it comes to commercialization of their work and overseeing derivate uses.[21] The novelty of this predicament has led to acute tensions, with apprehensions of inadequate testing of copyright law against the backdrop of AI training process, this leaves creators vulnerable, developers operating in uncertain legal waters. This essentially leads to uncertainty regarding whether mass computational use of copyrighted materials can be taken up as violation of intellectual property rights, especially copyrights.[22]

Research Objectives

1. To comprehensively examine legal frameworks and judicial decisions governing Fair Use (USA) and Fair Dealing (India) as applied to copyrighted works.

2. To analyze US based cases and judicial decisions on AI training and extent of copyright laws applicable.

3. To critically evaluate the efficiency of copyright laws to adequately cover the subject

[17] Sag, *supra* note 3, at 7–8; Charlesworth, *supra* note 8, at 7.
[18] Xia, *supra* note 6, at 3.
[19] 17 U.S.C. § 106(1) (2018); Sag, *supra* note 3, at 6.
[20] Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOUS. L. REV. 101, 103 (2025).
[21] Sag, *supra* note 3, at 4–5.
[22] U.S. COPYRIGHT OFFICE, COPYRIGHT AND ARTIFICIAL INTELLIGENCE: PART III – GENERATIVE AI TRAINING 3 (2025).

matter of AI training.

Research Questions

1. What is the doctrinal approach of Fair Use (USA) and Fair Dealing (India) and courts' interpretation of these principles?

2. How are US courts applying Fair Use test to AI Training on copyrighted works?

3. What is the impact of the Application of Fair Use and Fair Dealing Principles on AI developers across jurisdictions?

4. What is the extent of adequacy of copyright law to comprehensively address the challenges associated with exceptions in AI training?

Research Gaps

The primary literature available addresses fair use and AI Training in the jurisdictional limits of United States of America. Recent work carried out by Indian scholars have undertaken examining section 52 of the Copyright Act, 1957, however, there exists a critical gap in addressing a systematic comparative analysis between Fair Use and Fair dealing specifically applicable to Training AI Models. The current literature merely accounts for a descriptive examination of the present judicial framework and requires a thorough critical analysis. The current judicial framework is haphazard with each case bringing out varied outcomes, which limits drawing deeper analysis.

Doctrinal Framework in Copyright law

Fair Use: The US exception

The U.S. copyright Act under section 107 represents the cornerstone of copyright law's efforts to balance creator interests, rights with wider public interests concerning access and innovation.[23] Fair use is the seen as the exception and an equitable principle which permits unauthorized use or adaptation of a copyrighted work where certain conditions are met, this

[23] 17 U.S.C. § 107 (2018); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. (forthcoming 2024).

contrasts statutory exemptions that specify when infringement does not occur.[24] The doctrine of fair use has a rich historical background of precedents that began as a judicial interpretation of copyright frameworks in an attempt to understand when borrowed work relied heavily on an earlier existing work and seldom contributes of its originality, the same principles have now been equipped to address the growing concerns in the AI domain and technological challenges.[25]

The codification of fair use in the Copyright Act, 1976 in USA established a non-exhaustive four factor test that helps courts determine when a particular instance qualifies as fair use[26] and when it is an infringement. The four factors are:

(i) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(ii) the nature of the copyrighted work;

(iii) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

(iv) the effect of the use upon the potential market for or value of the copyrighted work[27]

The effect of using this framework is that courts now have substantial discretionary powers to evaluate the circumstances on a holistic manner, using factual matrix and contexts, rather than applying rigid legal rules to complex situations.[28]

The modern application of fair use focuses on 'transformative use' as propounded by Justice Pierre Leval as reinforced by the Supreme Court in *Campbell v. Acuff-Rose Music Inc.* (1994).[29] Transformative use employs the principle in a way where the use is fundamentally different from that of the source used or the use of the copyrighted material gives it an explicitly different expression.[30] This reflects copyright law's purpose of creating a balance

[24] Sag, *supra* note 1, at 8.
[25] *Id.*
[26] 17 U.S.C. § 107 (2018).
[27] *Id.*
[28] Sag, *supra* note 1, at 8–9
[29] Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1110–11 (1990); Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578–79 (1994).
[30] Leval, *supra* note 7, at 1111.

between promoting the creation and the distribution of new works, and interests of the authors.[31] The Supreme Court in the *Campbell* case asserts that copyright law permits transformative use, it enables referencing and interpretation of existing work, and thus fosters utilitarian purpose in promoting science and arts, the restriction on using copyrighted materials will inevitably hinder progress in various sectors.[32]

The *Andy Warhol Foundation v. Goldsmith* (2023) case clarified that fair use analysis must be taken up on a case-by-case basis rather than taking a holistic stand.[33] The courts analyzed whether Warhol Foundation's commercial licensing of color silkscreen prints utilizing photographer Lynn Goldsmith's 'Prince' portrait can be constituted as fair use, it ruled that it was not fair use, even though the foundation acknowledged modifications to the original photograph.[34] The rational behind it is that in case the secondary use exhibits the same function as the original work, where both offer commercial licensing, the secondary use will nor qualify as transformative use, merely because there was artistic modification; and this, it will not constitute fair use.[35]

Applying these developed principles on AI training will be analyzed in the upcoming sections.

Fair Dealing: The Indian exception

The Indian approach to copyright exception takes a diametrically varied stance compared to the US framework. The codification of 'fair dealing' rather than 'fair use' in section 52 of the Copyright Act, 1957 is where the difference in approach is clearly visible. Section 52 provides an exhaustive list of circumstances where an unauthorized use of copyrighted work will not constitute as infringement, this approach distinguishes itself from the equitable and open-ended principles adopted in the US Copyright Law.[36] Ultimately, fair dealing doctrine aims to harmonize and create a balance between the rights and economic interests of an artist and the broader requirements of accessibility to the society.[37]

---

[31] Campbell, 510 U.S. at 575.
[32] *Id.*
[33] Andy Warhol Found. v. Goldsmith, 143 S. Ct. 1258, 1284 (2023).
[34] *Id.* at 1267–71.
[35] *Id.* at 1284–85.
[36] Copyright Act, 1957, § 52, No. 14 of 1957, INDIA CODE (2024); Kailash Chauhan, *Generative AI, Text & Data Mining and the Fair Dealing Doctrine*, 30(1) J. INTELL. PROP. RTS. 77, 78 (2025).
[37] Chauhan, *supra* note 14, at 78.

Section 52 enumerates various categories of fair use for a wide variety of works including literary, dramatic, artistic, musical, computer programs for private use, research, educational purposes, criticism or review, reporting among other uses.[38] This section has procured an exhaustive list that helps navigate specific circumstances, and thus, sharply contrasts the US law.[39]

Indian courts have interpreted fair dealing provisions liberally given the rigid legal framework within which they operate. In *Wiley Eastern Ltd. v. Indian Institute of Management* (1996), the Delhi High Court established the purpose of Section 52 being protection of freedom enshrined in Article 19(1) of the Indian Constitution, the criteria in the case were to protect works of reporting current news, research and educational materials.[40] This has liberal approach backed by a constitutional grounding has enabled the courts expand the application of fair dealing in a way that encompasses broader concepts like fair use and the underlying principles of value of content utilized, purpose of use and competing likelihood between original and new work.[41]

Indian courts in deciding fair dealing have considered factors similar to those of fair use including purpose and character of the use, substantial difference between original and subsequent work, amount and proportion of similarity between the works and economic harm to owner's copyright market.[42] The courts have reflected on whether the unauthorized use was in bad faith or with improper motive, focusing on whether the user had knowledge if the work used was itself infringing; though courts have relied on the transformative quality and transformative purpose of use.[43]

The 2012 amendment to the Copyright Act was significant in that it expanded the scope of the section; it widened the provisions of the law to fit in societal needs.[44] This solidifies that fair dealing doctrine ought to evolve along with technological and social changes, however, it is also observed that these changes in law remain restricted by the categories enlisted in

[38] Copyright Act, 1957, § 52.
[39] Chauhan, *supra* note 14, at 78.
[40] Wiley E. Ltd. v. Indian Inst. of Mgmt., DLT 281, para. 19 (1996).
[41] Chauhan, *supra* note 14, at 79–80.
[42] *Id.* at 79.
[43] *Id.* at 79–80.
[44] Chauhan, *supra* note 14, at 79.

section 52 of the act.[45]

Foundational Precedents & Comparative Analysis between US and India

Landmark cases from both the US and India have shaped and defined principles that form a crucial party of contemporary copyright analysis. The *Google LLC v. Oracle America Inc.* (2021) addressed Google's copying of thousands of lines of Oracle's Java API to develop their platform is fair use.[46] The court ruled that the use of the source code in this case construed as transformative since it fundamentally served a different function which was to enable new software platforms rather than Oracle's original use of documenting java's API.[47] Although the use of copyrighted work in this case by Google was substantial in volume and amount, the material was utilized for an entirely different purpose technologically. This signifies that transformative analysis can allow space for large scale use of copyrighted materials when secondary use exhibits distinct objective.[48]

The Indian jurisprudence through *Civic Chandran v. Ammini Amma* (1996) and *University of Oxford v. Rameshwari Thakur &Anr.* (2009) has established that courts would apply fair dealing principle in a flexible manner beyond the express provision of law and the enlisted circumstances mentioned in section 52 to extend serve constitutional protection of freedom of expression while simultaneously making information more accessible.[49] These cases capture the perspective that courts have adopted a broader interpretation when matters concern social benefits, albeit the law limits such flexibility when compared to fair use principles enshrined in the US framework.

The fundamental difference between US and Indian provisions is that fair use and fair dealing both can lead to different outcomes for a similar factual matrix. Fair use doctrine concerns with a wholistic approach, "Does this use advance copyright's purpose even though unauthorized when all factors are weighed in?"[50] Fair dealing on the other hand deals with "Is this use applicable or allowed within the enumerated framework and does it matches up with conditions

[45] *Id.*
[46] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 28–29 (2021).
[47] *Id.*
[48] *Id.*
[49] Civic Chandran v. Ammini Amma, PTC 670 (Ker. H.C. 1996); Univ. of Oxford v. Narendra Publ'g House & Ors., CS(OS) 1656/2005 (Delhi High Court).
[50] Sag, *supra* note 1, at 10–11.

listed in the specific provision?"[51] The distinction is what helps the research draw the conclusion that depending on the principle applied, similar cases might see different outcomes. Further, this structural difference implies that a use of copyrighted work may satisfy or qualify as fair use in USA and fail to meet requirement of fair dealing in India if it does not relate precisely to the provisions of section 52, in essence creating a paradoxical situation.

AI Training and Copyright laws: A US Perspective

Training AI models in the US has seen an unprecedented use of copyrighted materials as source data, this has triggered the use of reproduction right under section 106 (1) of the US Copyright act.[52] Generative AI tools like "ChatGPT" "LLaMA", "Stable Diffusion" "DALL-E" are trained on billions of copyrighted as well as uncopyrighted data sources taken from the internet, open access sources such as books, news posts, articles, codes, images, music, social media posts etc.[53] Stable Diffusion has been trained on approximately 2 billion image sources from LAION 5B dataset with limited regard for copyright license, authorization or even consent of owners and creators.[54] Meta has a large language module LLaMA whoch was trained on publicly available data including resources from Gutenberg Project, Wikipedia, GitHub and other open sources.[55]

Using and copying that is unavoidable, inherent and necessary. Scholars have concluded that developers typically create a semi-permanent copy of the training data locally on servers for technical reasons, to avoid overfitting models, enable deduplication, bias analysis, filtering any toxic content and retraining the model.[56] Some scholars have theorized that the computational analysis of data is possible without creating the temporary or local copies, the practice is not common in present AI Development.[57] Scholars have further analyzed that this copying makes application of section 106 (1) crucial[58] and that it triggers 'colorable claims of copyright

[51] Chauhan, *supra* note 14, at 78–79.
[52] Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. (forthcoming 2024).
[53] *Id.* at 6; Edward Lee, *Fair Use and the Origin of AI Training*, 63 HOUS. L. REV. 101, 104–10 (2025).
[54] Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator* (Aug. 30, 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/.
[55] Hugo Touvron et al., *LLaMA: Open and Efficient Foundation Language Models*, arXiv preprint 2302.13971 (2023).
[56] Sag, *supra* note 1, at 7.
[57] *Id.*
[58] *Id.*

infringement' regardless whether training of AI models may constitute a fair use or not.[59]

Competing theories on Fair Use debate

The jurisprudence and scholarly debate in America addressing copyright infringement has taken two bifurcated paths, represented most clearly in the contemporary case laws and literatures speaking on 'transformative use' approach and the 'expressive exploitation' critique.

Transformative Purpose Framework

Ai training serves a fundamentally transformative purpose that is distinct from the original works intended purpose, thus, making it fair use.[60] The notion is that copyrighted works used to develop AI can be taken as comparable case to technology driven fair use cases, such as Google digitizing books for text-search functionality and even Sony's fortification of their reverse engineering for interoperability.[61] Emphasis has bee made the purpose of AI training, R&D has been technological innovation that is advancing US's standing on a global scale, this chalks up as distinct purpose and varied character from that of original works intended purpose and communicative expression.[62] The pivotal factor is here that copyrighted material act as raw data that is then processed by AI systems where it extracts statistical patterns and doesn't use the data as expressive content for derivative adaptation or reproduction.[63]

Lastly, the Copyright Act's progress clause enables courts to weigh technological progress equally with creative progress in fair use debated.[64] Copyright aims to promote science and useful arts, scientific and technological innovation.[65] Thus, it is contented that courts should use this principle in their application of fair use factors, especially when assessing the economic harm or copyright of creators outweigh technological purpose that is transformative in nature beneficial to the public at large.

Expressive Exploitation Criticism

On the other hand of transformative purpose is a critique that contests the characterization of

[59] *Id.* at 6
[60] Lee, *supra* note 1, at 103–60.
[61] *Id.* at 127–45.
[62] *Id.* at 147–60.
[63] Sag, *supra* note 1, at 13–24.
[64] Lee, *supra* note 1, at 143–45; U.S. CONST. art. I, § 8, cl. 8.
[65] Lee, *supra* note 1, at 143–45.

AI training as non-expressive use, by challenging that AI models' training is exploitative of the copyrighted works used as data.[66] It is contented that models like GPT-4 do not 'discard' training data after coding the statics properties, rather it is broken into tokens that are relied on for longer durations, this process is exploitative of the expressive elements of copyrighted works used as training data.[67] This use of data as token for generating outputs can be seen as ongoing exploitation.[68] This is how the argument states that AI training differs from precedented cases of fair use like Google Booka, HathiTrust where the copying served a functional role and not as an expressive end.[69]

Contemporary cases

1. ***Bartz v. Anthropic PBC (June 2025)***

In this case, the federal judge ruled that training AI Model on copyrighted works constitutes as fair use.[70] It was alleged that Anthropic trained Claude AI on copyrighted books without any authorization, which lead to the petitioner seeking damages and injunction.

The Judge applied the four-factor test enshrined in section 107 of the copyright act in a methodical manner. The court found that the use was "spectacularly transformative" with regards to subsection1 inspecting the purpose and character of use.[71] The court went on to emphasize that training serves the purpose of developing a new technology where AI model is capable of performing analytical and generative actions, this does not necessarily substitute original works' expressive use or intended communication.[72] The rational in this case that the books were used as data, not for replication or adaptation.

The judge when assessing Subsection 4 addressing the economic and market harm, rejected market dilution theory saying it is unsupported by precedents. It was found that there is little to no competition between original works created by humans and AI-generated works, this it does not cause any cognizable harm under the copyright law. Exception to this general notion

[66] Jacqueline C. Charlesworth, *Generative AI's Illusory Case for Fair Use* 4–16 (2024).
[67] *Id.* at 7.
[68] *Id.*
[69] *Id.* at 10–12.
[70] Bartz v. Anthropic PBC, No. 3:24-cv-05417, 2025 WL 1741691 (N.D. Cal. June 23, 2025).
[71] *Id.*
[72] *Id.*

is that if AI output directly can be directly substituted for specific copyrighted works.[73]

The court thus concluded that AI systems might be capable of eventually displacing human writing tasks, the general competitive effect differ from market substitution addressed by copyright act and fair use principles.[74]

2. *Kadrey v. Meta Platforms Inc. (June 2025)*

In another judgement delivered shortly after the Bartz case, the California Northern District Court found that Meta's training of its AI model on copyrighted books was transformative and thus, fair use.[75] The court applied the reasonings that were parallel to the Bartz case, where technological advancement through AI represents distinct purpose from that of original work. Similarly, market harm analysis assessment also followed an analogous application ad interpretation of the law.

The congruence of the *Bartz* and *Kadery* cases, though decided by different judges suggests emerging trends and consensus in decision making concerning AI models' training. Especially in context of research and development purpose rather than direct competitive substitution.

3. *Thomson Reuters Enter. Centre GmbH v. ROSS Intelligence Inc.* (February 2025)

Preceding the cases mentioned above, in this case the Delaware court issued a contrary ruling. It rejected fair use in AI training entirely.[76] It was alleged that ROSS Intelligence has used and copied Westlaw headnotes to train it AI system designed for providing legal research analysis to users, which competes directly with Reuters research products directly.

The case's focus was distinguished from technology based fair use by focusing on market substitution. It was found that ROSS's AI output competed directly in market with Reuters for legal research analysis, where the latter is a key player.[77] Unlike the Google Books issue, here the AI model was in direct competition and used the copyrighted data in a non-transformative

[73] *Id.*
[74] *Id.*
[75] Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417, 2025 WL 1752484 (N.D. Cal. June 25, 2025).
[76] Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc., No. 1:20-cv-00613, 2025 WL 456789 (D. Del. Feb. 11, 2025).
[77] *Id.*

manner and non-expressive use was noted. Hence, the use of copyrighted material in this case is not fair use and is in fact copyright infringement.

Reconciling conflicting position

The divergent outcomes in the US jurisprudence can be explained by segregating the phases of training of AI and deployment (output generation) of AI. The courts have regarded copying of technology for development, where work is used as an input to be fair use, but when the output is a direct result of copied work, it harms the original works' rights and is not fair use.

India's Copyright Laws applicable to AI training

Section 52 of the Copyright Act, 1957 was formulated long before the AI boom, it operated in an era where research was based on print media, books and photocopying materials for reference.[78] The framework of section 52 captures the parliamentary intent to precisely codify and enlist cases/circumstances that will not construe as copyright infringement such as journalism, private use, education, research and activities not for profit among others.[79] Conversely, the US fair use doctrine is flexible, without categorical limitations and address each case from a contextual analysis.[80] In India, protection of creators and authors rights have been prioritized through concrete provisions and exceptions, rather than judicial discretion.

In *Wiley Eastern Ltd. v. IIM* courts have adopted a wider scope of interpretation going beyond the statutory language in an attempt to protect constitutional freedoms.[81] Subsequent cases and decisions have centered judicial approach towards an open and flexible one comparable to the fair use principle, even when acting within the constrained limits of the Indian fair dealing law.[82] This degree of flexibility was especially beneficial during 2010s Information technology developments, but proves to be inadequate for the AI era and addressing the large scale operation of AI models' training.

[78] Copyright Act, 1957, § 52, No. 14 of 1957, INDIA CODE (2024); Kailash Chauhan, *Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime*, 30(1) J. INTELL. PROP. RTS. 77, 78 (2025).
[79] *Id.*
[80] 17 U.S.C. § 107 (2018); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. (forthcoming 2024).
[81] Wiley E. Ltd. v. Indian Inst. of Mgmt., DLT 281, para. 19 (1996).
[82] Civic Chandran v. C. Ammini Amma, PTC 670 (Ker. H.C. 1996); Chauhan, *supra* note 1, at 79–80.

Doctrinal Mismatch Created by AI

When the provisions of section 52 are applied to the regime of generative AI, fundamental and deep-rooted structural challenges come to the surface. The primary provision of the section, 52(1)(a) which allows fair dealing for "personal and private use; including research" was drafted prior to the AI revolution, it lacks considerations towards commercial corporate training, mass scale computation of copyrighted works, non-expressive use where works are treated as data rather than expressive content.[83]

Commercial AI creators and developers operate for profit, that generate large sums of revenue from these AI models that are trained on copyrighted materials. Private use mentioned in section 52(1)(a)(i) effectively barricades this commercial exploitation, thus creating a mismatch.[84] Unlike US doctrine of fair use, that lacks any distinction for commercial and non-commercial use, Indian concept of fair dealing implies that commercial activity fall outside the purview of this exception. However, the challenge arises when this use is excluded from any relaxations. Doing so will ultimately prevent AI developers to effectively train their models, increasing reliance on international systems and diminishing economic boost of our economy. This also indirectly hampers creators' economic benefits by limiting their licensing agreements with Indian developers.

The fair dealing exemption allows for transient or incidental storage for transmission purpose, however, AI training requires a local server for storing copyrighted works for technical purposes, deduplication, bias analysis among other processing requirements.[85] While storage under section 52 is provided for private use, it lacks adequate coverage on storage for computational extraction large datasets simultaneously.[86]

Indian courts have not adequately defined whether research encompasses commercial AI development. It has been argued by scholars that it is critical to identify whether the commercial data mining is construed research or not under section 52.[87] US precedents have clarified that commercial technological development constitutes as fair use in the *Google v. Oracle* case.[88]

[83] Chauhan, *supra* note 3, at 78–79.
[84] *Id.* at 79.
[85] Sag, *supra* note 1, at 7.
[86] Copyright Act, 1957, § 52(1)(a)(ii); Sag, *supra* note 3, at 7.
[87] Chauhan, *supra* note 3, at 79.
[88] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 28–29 (2021).

Indian fair dealing doctrine has not addressed this yet, leaving creators and companies in a dilemma.

Application of fair dealing

The application of section 52 has proven to be difficult in the Indian context, given the structural rigidity of the law, a liberal interpretation will allow the AI companies to keep carrying business at a risk of creator rights, contrary to this, when the interpretation is stricter, the AI companies will simply move away from Indian sources, halting licensing, curbing Indian AI development and brining us at a crossroads.[89]

The ANI v. OpenAI case

Indian legal framework has directly addressed the dilemma of AI training on copyrighted materials in the case filed by Asian News Internation against OpenAI. The case was filed in Delhi HC in December 2024.[90] ANI claims that OpenAI has violated section 14 of copyright act by copying a large volume of ANI's resources to train ChatGPT without due authorization or consent or any other agreement or compensation.[91] OpenAI has in its defence claimed reliance on section 52, arguing that the training constitutes as fair dealing as it serves transformative purpose, despite commercial deployment.

The case is presently pending before the court but the primary concern here lies in addressing the following legal issue: Is computational analysis research? The decision will help clarify a largely debated legal standpoint, whether section 52 encompasses commercial use of copyright materials in training AI models as fair dealing or not.[92] It will also address if open access and internet sources will require permission from creators prior to use, or will come as fair dealing exemption.[93]

The contemporary legal scenario in India is rather murky, even with flexibility and judicial discretion, accommodating AI Training in the rigid framework of section 52 will require

[89] Chauhan, *supra* note 1, at 78–79.
[90] ANI Media Pvt. Ltd. v. OpenAI Inc., CS(COMM) 1028/2024 (Delhi High Court, filed Dec. 2024).
[91] Sapna Singh, *Feeding the Algorithm: Legal Challenges in AI Training Data for Educational Progress* 2 (2024).
[92] *Id.* at 2–3.
[93] *Id.*

legislative intervention.[94]

Conclusion & Suggestions

This research has reflectively showcased that section 52 and Indian fair dealing provisions are at present unsuited to address the large-scale AI revolution. While US is better equipped, India requires a legislative overhaul in its laws to better address these emerging concerns.

Though it is noted in scholarly literature that AI training is fair use, its practical implications pose risks such as 'memorization' of data. Thus, even at a global scale, the laws need to be refreshed to fully capture the nuances created in this legal0tchnological landscape.

The way forward looks promising if the statutory exceptions will add a targeted clause for AI related matters, specifically AI training. Removal of commercial barriers and restrictive interpretation of the laws will allow for a more holistic protection regime.

Companies should also be liable in adopting transparent practices and acknowledge steps taken to ensure data security and transparency. They can also adopt a compensatory practice where creators are duly credited and given fair compensation for the use of their works.

Creating a balanced ecosystem will empower creators, enable responsible development of AI and remove legal uncertainty.

[94] Chauhan, *supra* note 3, at 78.