
THE COPYRIGHT IMPLICATIONS OF GENERATIVE AI TRAINING: THE NEED FOR LEGAL REFORMS IN INDIA

S. Rahamathunisa, The School of Excellence in Law, The Tamil Nadu Dr. Ambedkar Law University, Chennai

ABSTRACT

Generative Artificial Intelligence has posed serious problems for copyright law with regard to the use of copyrighted resources for AI training. AI trained using the large datasets of book, articles, music, sounds and other protected works without the consent of the copyright owners. This study examines whether such AI training is protected by fair dealing as per section 52 of the Copyright Act of 1957 or Copyright infringement. This paper insists AI training represents a large-scale extraction of the creative and financial value contained in copyrighted works rather than just a technological learning process. Generative AI systems industrially process and profitably use copyrighted expression to produce marketable outputs, in contrast to conventional human learning. This paper also examines that fair dealing is inadequate to regulate the AI training. The study underscores the need for statutory reform and finds a legal gap in the Copyright Act of 1957 regarding the use of AI datasets. The study proposes the implementation of AI-specific copyright laws, dataset disclosure requirements, and licensing mechanisms in order to achieve a balance between technological innovation and creative protection. It also conducted a comparison analysis with the US and EU concerning fair use and text and data mining. The study highlights a legal loophole in Indian copyright law pertaining to the usage of AI datasets and emphasizes the necessity of legislation reform.

Keywords: Generative Artificial Intelligence, Copyright, Fair dealing, Copyright infringement, AI Training, Fair use.

Introduction

The development of generative AI has revolutionized the production and distribution of digital material. AI programs like ChatGPT and Midjourney may produce literature, images, music, and other creative outputs that resemble those of a human. Large-scale datasets of songs, music, novels, and other protected works are used to train this system. There are serious ethical and legal issues with the growing use of copyrighted content in AI training. Without the copyright owner's permission, AI developers use copyrighted content for AI training. The AI uses the expression, style, and patterns of the copyrighted works when producing its outputs. The outputs generated by AI systems can compete commercially with the original authors whose works were used for training. The original purpose of copyright law was to control how people used creative works. The Fair dealing doctrine allows for the restricted use of copyrighted information for private study, research, criticism, and reviews. However, unlike traditional human learning, generative AI extracts large-scale datasets for commercial purposes.

This paper contends that AI training is an extraction of creative and commercial value from copyrighted works rather than just technological learning. The paper analyses the insufficiency of the Copyright Act of 1957 in addressing AI-driven technology and examines whether AI training can lawfully come under the purview of fair dealing under Indian law.

Understanding AI Training

The process of learning patterns, creative styles, information, images, sounds, and linguistic structures from the available data is known as AI training. A dataset is the collection of information used to train AI.

Massive collections of research articles, news, stories, books, music, blogs, sounds, and artwork and other copyrighted work are used to train generative AI systems like ChatGPT, Gemini, and Midjourney. To find patterns, writing styles, frameworks, and creative processes, the AI system continuously examines this data. AI responds to prompt with output that resembles the patterns seen in the previously trained copyrighted works.

Extraction of creative value

The process by which AI systems profitably use the intellectual labour contained in copyrighted works without paying the original creators is known as "extraction of creative value." Because

trained AI systems produce commercial outputs, the value obtained from protected information becomes economically advantageous for AI developers. Therefore, AI training is a methodical technical exploitation of copyrighted creativity rather than just passive learning.

Fair Dealing in the Copyright Act of 1957

A legal basis for protecting authors' rights to their original works is provided under the Indian Copyright Act of 1957. Although copyright infringement is defined under Section 51 of the Copyright Act of 1957, certain uses of copyrighted works are excluded from infringement under Section 52. The term "fair dealing" is frequently used to describe these exceptions. The objective of the fair dealing is to maintain a balance between the exclusive rights of copyright owners and the public interest in accessing copyrighted works for socially beneficial purposes such as education, research, commentary, and expression. The definition of fair dealing emanates from court interpretation and is not found in the Copyright Act, 1957.

The scope of fair dealing in the Copyright Act of 1957 is limited to the comprehensive list in section 52 in contrast to the fair use concept in the United States, which permits a four-factor analysis that evaluates the purpose of use, the nature of the original work, the amount you copied, and its market effect on the creator's earnings. The fair dealing in section 52 is rigid and it is difficult to address new technology advancements because of this inflexible structure.

Large datasets of books, papers, artwork, music, and other copyrighted content are used to train AI models. Large-scale automated copying and computational analysis of artistic creations are part of this process. According to this study, AI training encompasses more than just "use" or "learning" in the conventional sense. Rather, it is a methodical process of extracting creative and economic value from copyrighted works, wherein the authors' expressive labour is converted into economically lucrative outputs produced by AI systems.

The conventional understanding of fair dealing, which was first intended for human-centric activities like private study, criticism, or reporting on current events, does not apply to this new kind of large-scale value extraction. Machine-based, large-scale processing of copyrighted materials for commercial model training was not anticipated by the Section 52 notion of fair dealing. India's current legal framework is insufficient to regulate AI training activities.

Legality of using copyrighted works in AI Training

Whether using copyrighted content for AI training is fair dealing or copyright infringement is

at the core of the legal issue.

A crucial legal question lies at the centre of this debate: Should AI training be seen as ordinary learning similar to human learning, or does it take into account the unlawful extraction and commercial exploitation of copyrighted creative expression, which leads to copyright infringement?

Copyright Infringement and Extraction of Creative Value

On the one hand, AI training goes beyond ordinary learning and violates copyright by taking creative and commercial value from works that are protected by copyright. AI training is different from ordinary human learning since it occurs on a large industrial scale. Human learning is intrinsically limited since individuals read certain sources, understand ideas, and acquire their own knowledge. Human learning usually does not replicate or profitably exploit vast quantities of copyrighted materials. Generative AI systems will extract patterns, styles, structures, and artistic processes from massive databases, assess millions of copyrighted works at the same time, and generate profitable results. Because AI training entails the industrial-scale computational extraction of creative expression, it is fundamentally different from traditional human learning. AI won't have human understanding in this scenario. The goal is to transform artistic expression into computationally useful data that can be sold, rather than just comprehending information.

As a result, straight copying is not the only problem. The commercial exploitation of creative labour without creators' knowledge, consent, or payment is the bigger issue. Furthermore, copyrighted information must frequently be stored, reproduced, and processed within datasets and computational systems as part of the training process. The exclusive rights conferred by the Copyright Act of 1957, especially the rights to reproduction and adaptation, may be interfered with by such activities.

Supporting AI Training under Copyright law

AI developers contend that since AI systems primarily use copyright materials to learn patterns, relationships, and structures rather than directly recreating expressive content, AI training should not be automatically viewed as copyright infringement. Consequently, rather than producing exact replicas of protected works, generative AI systems produce new and

distinctive results. They further contend that factual data, concepts, and publicly accessible online content are often processed by AI systems and are typically not separately protected under copyright law. Original expression is protected by copyright law, not concepts, facts, or informational data. Thus, it shouldn't always be considered infringement to learn patterns or relationships from publicly accessible materials.

Moreover, artificial intelligence developers claim that stringent copyright limits on AI training could hinder technical innovation and digital development, necessitating some flexibility in copyright law to foster innovation. They further contend that AI systems do not purposefully "copy" copyrighted expression for direct replication; rather, they only analyse data in a technologically advanced way. There is still legal challenge because the Copyright Act of 1957 is not created for the machine learning technologies.

Copyright concerns in Deepfake Technology

Deepfake technology is one modern instance of possible copyright infringement. AI is used in deepfakes to alter realistic images, sounds, and movies that may appear to be identical to the original human-made work. The Deepfake system creates copies by analysing copyrighted works, including images, videos, and sounds.

Through continuous machine learning processes, AI systems can produce content that is difficult to distinguish from authentic works. The growing prevalence of deepfake technology presents significant ethical and legal issues because during training, a lot of copyrighted materials are used, authors might not agree to such use, AI-generated content might compete with original works in the market and Deepfake-generated content is currently subject to minimal legal regulation.

This deepfake technology shows how generative AI systems use algorithmic extraction and reproduction to profitably exploit artistic expression. Legal ambiguity about responsibility, author protection, and financial harm from AI-generated content is further exacerbated by the lack of explicit copyright laws relating to AI.

Doctrine of Transformative Use in Copyright Law

The concept of transformative use is important in today's copyright dispute, especially given the quickly evolving field of artificial intelligence. In generative AI, transformative use is

essential. According to copyright law, transformative use occurs when a new work adds significantly new expression, a new message, or a new purpose, changing the previous work's character and purpose. This idea became more significant during the ANI and OpenAI dispute.¹

In *Digital Collectibles Pte Ltd and Ors. v. Galactus Funware Technology Pvt Ltd and Anr.*², the Delhi High Court recently addressed the transformative use. As stated by the court, a use is transformative if it adds a new expression, character, or purpose to the original work.

Judicial Interpretation of AI Training

The legality of the AI training using the copyrighted work is still in dispute between the OpenAI and ANI.³ Uncertainty in the application of the Copyright Act of 1957 stems from the lack of clear judicial precedents and the AI-specific copyright provision.

Regarding the usage of datasets for AI, there are no appropriate legal provisions. The AI Works has not yet been determined by the courts and there are no fixed rules to follow for AI training.

Comparative Analysis : The US and EU Perspectives

Let's compare the Indian legal system to that of the US and the EU to gain a broad understanding of AI and copyright. In contrast to fair dealing in India, the US follows the doctrine of fair use, which is outlined in section 107 of the Copyright Act. The US fair use analysis involves a four-factor test that considers the purpose and character of the use, including whether it is transformative and of a commercial nature; the nature of the copyrighted work; the amount and substantiality of the portion used; and the effect of the use upon the potential market for or value of the copyrighted work⁴. India's fair dealing doctrine is more limited and less flexible when it comes to the new challenges that artificial intelligence presents.

The Supreme Court defined transformative use in *Campbell v. Acuff-Rose Music, Inc.*⁵, The Court held that the critical question is whether the new work “adds something new, with a further purpose or different character, altering the first with new expression, meaning, or

¹ ANI Media Pvt. Ltd. v. OpenAI OpCo LLC ,CS(COMM) 1028/2024

² 2023 SCC OnLine Del 2306

³ ANI Media Pvt. Ltd. v. OpenAI OpCo LLC ,CS(COMM) 1028/2024

⁴ Sandeep Kanak Rathod, Fair Use: Comparing US and Indian Copyright Law, JURIST (May 28, 2012), <https://www.jurist.org/commentary/2012/05/sandeep-kanak-rathod-copyright/>.

⁵ 510 U.S. 569 (1994)

message.”⁶

A legislative approach for text and data mining(TDM) has been established by the European Union. If libraries and research institutions have legal access to copyrighted material, they may use it for scientific research. Unless the copyright owner expressly declines permission, others, including businesses, may likewise use content for TDM. By using notices like machine-readable metadata, copyright owner can prevent such use(opt-out" procedure). As a result, the EU has a more defined legal system than India, whose regulations are still unclear.

Findings

The study finds that AI learning is really different, from human learning. This is because machine learning relies on automated large-scale data extraction. AI systems extract creative value and economic value from copyrighted works without consent from copyright owners. The fair dealing provisions under the Copyright Act of 1957 were designed for purposes like research, criticism, review, and private study, not for AI training or machine learning. India's fair dealing doctrine were designed for purpose like research, criticism, review. However ,the scope of fair dealing in context of AI generated work is unclear. The Copyright Act of 1957 is inadequate to address legal issues created by generative AI technologies. The use of copyright law to AI-generated content and AI training methods has problems with legal ambiguity, and legislative changes are desperately needed to strike a balance between technical advancement and copyrighted work owners' rights.

Recommendation

The Copyright Act of 1957 needs to be amended and it should have provisions, about AI training. These provisions should cover how generative AI systems use copyrighted works. The Governments need to introduce separate legal framework for AI-generated works. The Fair dealing doctrine in India need to updated for addressing AI training and AI-generated content. Licensing mechanism needs to introduce for the use of copyrighted works in AI training so that the copyright owner entitles to royalties and fair compensation. Transparency obligation needs to impose on the AI developer through statutory provisions such as disclose the copyright materials, training datasets, and data collection methods used for training AI models. The regulatory body need to introduce to oversee the AI training to protect the rights

⁶ Campbell v. Acuff-Rose Music, Inc, 510 U.S. 569 (1994)

of the copyright owners. India may adhere to the US fair use approach by enacting flexible exceptions that permit limited use of copyrighted works for innovation, research, and technological advancement while still respecting rights of owners. The European Union's Text and Data Mining (TDM) exceptions, which permit AI training on legally accessible content subject to protections and rights reservations by copyright holders, may also be adopted by India. The owners of copyright may specifically protect their rights against the use of their works for AI training through a fair "opt-out" procedure, similar to the EU model. All the Legal reforms should bring to balance technological innovation with the copyright holder's rights

Conclusion

The Copyright Act of 1957 is having difficulty with generative AI, particularly when it comes to using copyrighted works for AI training. AI training differs from conventional human learning in that it makes use of copyrighted works, depriving the owners of their creativity and financial benefit. Even though using copyrighted content to train AI would not necessarily be seen as copyright infringement, there are still legal issues when it causes financial or reputational damage. The Copyright Act of 1957 is insufficient to address machine learning, thus updated legislation is urgently needed to strike a balance between the protection of copyright owners' rights and technological innovation.

REFERENCES:

- [1] Kailash Chauhan, “Generative AI, Text & Data Mining and the Fair Dealing Doctrine:Examining the New Problem with the Old Regime” 30 Journal of Intellectual Property Rights(2024)
- [2] Dr.Sonali S.Gholve, Adv.Sagar K.Gholve, “Study of whether using Copyrighted Content for Training AI Models a Copyright Infringement?” 1 IJHCE(2025)
- [3] Tarun Kumar Singh Jayant & MS. Aradhana Yadav, “Generative AI and Fair Dealing under Indian Copyright Law” 5 IJLR(2025)
- [4] Ritesh Karale & Ritaja Chattopadhyay, “The Copyright Play: AI, Section 52, and the Acts of Fair Dealing” available at: <https://forum.nls.ac.in/ijlt-blog-post/the-copyright-play-ai-section-52-and-the-acts-of-fair-dealing/>(last visited on 25 Mar,2026)
- [5] Moomal Joshi ,“Web Scraping and AI Training Data: Copyright Challenges in the Age of Generative AI” available at: <https://www.khuranaandkhurana.com/web-scraping-and-ai-training-data-copyright-challenges-in-the-age-of-generative-ai> (last visited on 1 May,2026)
- [6] Lexport, Copyright in the age of AI: A case comment on ANI vs. OpenAI, CS(COMM)1028/2024availableat:[https://www.lexport.in/images/1748945827_Article%20I%20COPYRIGHT%20IN%20THE%20AGE%20OF%20AI%20A%20CASE%20COMMENT%20ON%20ANI%20VS.%20OPENAI%20CS\(COMM\)%201028-2024.pdf](https://www.lexport.in/images/1748945827_Article%20I%20COPYRIGHT%20IN%20THE%20AGE%20OF%20AI%20A%20CASE%20COMMENT%20ON%20ANI%20VS.%20OPENAI%20CS(COMM)%201028-2024.pdf) (last visited on 1 Apr,2026)