



# INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS

Open Access, Refereed Journal Multi Disciplinary  
Peer Reviewed Edition :

[www.ijlra.com](http://www.ijlra.com)

## **DISCLAIMER**

No part of this publication may be reproduced or copied in any form by any means without prior written permission of Managing Editor of IJLRA. The views expressed in this publication are purely personal opinions of the authors and do not reflect the views of the Editorial Team of IJLRA.

Though every effort has been made to ensure that the information in Volume 2 Issue 7 is accurate and appropriately cited/referenced, neither the Editorial Board nor IJLRA shall be held liable or responsible in any manner whatsoever for any consequences for any action taken by anyone on the basis of information in the Journal.

Copyright © International Journal for Legal Research & Analysis

IJLRA

## **EDITORIAL TEAM**

### **EDITORS**



### **Megha Middha**

*Megha Middha, Assistant Professor of Law in Mody University of Science and Technology, Lakshmangarh, Sikar*

*Megha Middha, is working as an Assistant Professor of Law in Mody University of Science and Technology, Lakshmangarh, Sikar (Rajasthan). She has an experience in the teaching of almost 3 years. She has completed her graduation in BBA LL.B (H) from Amity University, Rajasthan (Gold Medalist) and did her post-graduation (LL.M in Business Laws) from NLSIU, Bengaluru. Currently, she is enrolled in a Ph.D. course in the Department of Law at Mohanlal Sukhadia University, Udaipur (Rajasthan). She wishes to excel in academics and research and contribute as much as she can to society. Through her interactions with the students, she tries to inculcate a sense of deep thinking power in her students and enlighten and guide them to the fact how they can bring a change to the society*

### **Dr. Samrat Datta**

*Dr. Samrat Datta Seedling School of Law and Governance, Jaipur National University, Jaipur. Dr. Samrat Datta is currently associated with Seedling School of Law and Governance, Jaipur National University, Jaipur. Dr. Datta has completed his graduation i.e., B.A.LL.B. from Law College Dehradun, Hemvati Nandan Bahuguna Garhwal University, Srinagar, Uttarakhand. He is an alumnus of KIIT University, Bhubaneswar where he pursued his post-graduation (LL.M.) in Criminal Law and subsequently completed his Ph.D. in Police Law and Information Technology from the Pacific Academy of Higher Education and Research University, Udaipur in 2020. His area of interest and research is Criminal and Police Law. Dr. Datta has a teaching experience of 7 years in various law schools across North India and has held administrative positions like Academic Coordinator, Centre Superintendent for Examinations, Deputy Controller of Examinations, Member of the Proctorial Board*



## Dr. Namita Jain



*Head & Associate Professor*

*School of Law, JECRC University, Jaipur Ph.D. (Commercial Law) LL.M., UGC -NET Post Graduation Diploma in Taxation law and Practice, Bachelor of Commerce.*

*Teaching Experience: 12 years, AWARDS AND RECOGNITION of Dr. Namita Jain are - ICF Global Excellence Award 2020 in the category of educationalist by I Can Foundation, India. India Women Empowerment Award in the category of "Emerging Excellence in Academics by Prime Time & Utkrisht Bharat Foundation, New Delhi.(2020). Conferred in FL Book of Top 21 Record Holders in the category of education by Fashion Lifestyle Magazine, New Delhi. (2020). Certificate of Appreciation for organizing and managing the Professional Development Training Program on IPR in Collaboration with Trade Innovations Services, Jaipur on March 14th, 2019*

## Mrs.S.Kalpana

*Assistant professor of Law*

*Mrs.S.Kalpana, presently Assistant professor of Law, VelTech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Avadi. Formerly Assistant professor of Law, Vels University in the year 2019 to 2020, Worked as Guest Faculty, Chennai Dr. Ambedkar Law College, Pudupakkam. Published one book. Published 8 Articles in various reputed Law Journals. Conducted 1 Moot court competition and participated in nearly 80 National and International seminars and webinars conducted on various subjects of Law. Did ML in Criminal Law and Criminal Justice Administration. 10 paper presentations in various National and International seminars. Attended more than 10 FDP programs. Ph.D. in Law pursuing.*



## Avinash Kumar



*Avinash Kumar has completed his Ph.D. in International Investment Law from the Dept. of Law & Governance, Central University of South Bihar. His research work is on "International Investment Agreement and State's right to regulate Foreign Investment." He qualified UGC-NET and has been selected for the prestigious ICSSR Doctoral Fellowship. He is an alumnus of the Faculty of Law, University of Delhi. Formerly he has been elected as Students Union President of Law Centre-1, University of Delhi. Moreover, he completed his LL.M. from the University of Delhi (2014-16), dissertation on "Cross-border Merger & Acquisition"; LL.B. from the University of Delhi (2011-14), and B.A. (Hons.) from Maharaja Agrasen College, University of Delhi. He has also obtained P.G. Diploma in IPR from the Indian Society of International Law, New Delhi. He has qualified UGC - NET examination and has been awarded ICSSR - Doctoral Fellowship. He has published six-plus articles and presented 9 plus papers in national and international seminars/conferences. He participated in several workshops on research methodology and teaching and learning.*

## **ABOUT US**

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH & ANALYSIS

ISSN

2582-6433 is an Online Journal is Monthly, Peer Review, Academic Journal, Published online, that seeks to provide an interactive platform for the publication of Short Articles, Long Articles, Book Review, Case Comments, Research Papers, Essay in the field of Law & Multidisciplinary issue. Our aim is to upgrade the level of interaction and discourse about contemporary issues of law. We are eager to become a highly cited academic publication, through quality contributions from students, academics, professionals from the industry, the bar and the bench. INTERNATIONAL JOURNAL FOR LEGAL RESEARCH & ANALYSIS ISSN 2582-6433 welcomes contributions from all legal branches, as long as the work is original, unpublished and is in consonance with the submission guidelines.

# **COPYRIGHT VS. INNOVATION: NAVIGATING FAIR USE FOR AI TRAINING DATA**

AUTHORED BY - SHRIYASHA KHANDIGE

## **Abstract:**

The development of artificial intelligence (AI) hinges on massive datasets for training purposes. This raises concerns regarding copyright infringement when copyrighted works are included in the training data. This abstract explores the concept of fair use as a potential defence in such scenarios.

The analysis highlights the ongoing debate surrounding fair use and AI training. While some argue that the transformative nature of AI development qualifies as fair use, others express concerns about the potential harm to copyright holders. The abstract examines key considerations within the fair use framework, including the purpose and character of the use, the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect of the use upon the potential market.

This research is based on the US jurisdiction and its development because of the comparative evolve in the jurisprudence compared to the rest of the world.

Recent cases and ongoing discussions are explored to provide a nuanced perspective on the evolving legal landscape. The abstract concludes by emphasising the need for potential solutions, such as clearer guidelines or licensing models, to ensure the responsible development of AI while protecting intellectual property rights.

## **Introduction:**

The remarkable advancements in artificial intelligence (AI) have revolutionized numerous fields, from healthcare and finance to creative industries. However, this progress hinges on a crucial first step: training AI models on vast amounts of data. This data often includes copyrighted works, such as text, images, and music, raising a critical question: does using copyrighted material for AI training constitute copyright infringement?

This paper delves into the complex intersection of intellectual property law and AI development, with a specific focus on the concept of fair use. Fair use is a legal doctrine that permits limited use of copyrighted material without the copyright holder's permission for purposes such as

criticism, commentary, or news reporting. However, its application to AI training remains an area of ongoing debate.

This paper explores the arguments for and against considering AI training as fair use. Proponents highlight the transformative nature of AI, arguing that training data is merely a tool for creating entirely new and innovative outputs. Conversely, some copyright holders express concerns about the potential for AI to supplant their works or devalue their market.

By examining the four-factor fair use test – purpose and character of the use, nature of the copyrighted work, amount and substantiality of the portion used, and the effect of the use upon the potential market – this paper analyzes the legal viability of using copyrighted material for AI training. We will explore relevant case studies and emerging legal frameworks to understand how courts are currently grappling with this issue.

Ultimately, this paper aims to provide a comprehensive understanding of the fair use debate in the context of AI training. By navigating the complex legal landscape and exploring potential solutions, we hope to foster a dialogue that promotes innovation in the AI field while safeguarding the rights of creators.

#### Arguments for Fair Use in AI Training

Proponents of fair use in AI training highlight several key arguments. Firstly, they emphasize the transformative nature of AI. Unlike traditional copying, training data is not used to create derivative works or compete directly with the copyrighted material. Instead, it serves as a building block for entirely new and innovative outputs. AI models, once trained, can generate novel content, translate languages with exceptional accuracy, or identify patterns unseen by the human eye.

Secondly, proponents argue that the amount and substantiality of copyrighted material used in training is often minimal compared to the overall dataset. AI models are typically trained on massive datasets encompassing millions or even billions of data points. The copyrighted material might constitute only a small fraction of this data, often serving as a reference point for the model to learn underlying patterns and relationships.

Thirdly, supporters of fair use contend that AI training has a positive impact on creativity and innovation. By providing researchers and developers access to training data, fair use fosters the advancement of AI technology, which in turn can be used to create new tools for creative expression. For instance, AI can generate original musical compositions or artistic styles inspired by existing works but ultimately distinct from them.

## Arguments Against Fair Use in AI Training

Opponents of fair use in AI training raise concerns about the potential negative impact on copyright holders. They argue that the sheer scale of training data utilized by large corporations could have a detrimental effect on the market value of copyrighted works. If AI models can readily replicate the style and content of existing works, there's a risk that the demand for original creations diminishes.

Furthermore, some copyright holders express anxieties about the lack of transparency in AI training algorithms. The specific ways copyrighted material is used within the training process can be opaque, making it difficult to assess the potential harm to their works.

Finally, opponents caution against inadvertently granting a "blank check" to AI developers. Without clear guidelines or limitations on fair use for AI training, copyright holders might find themselves unable to protect their works from unauthorized commercial exploitation.

## Applying the Fair Use Test

The legal viability of using copyrighted material for AI training hinges on the four-factor fair use test established in the United States Supreme Court case *Campbell v. Acuff-Rose Music* (1994)<sup>1</sup>.

This test considers:

1. The purpose and character of the use: Is the use transformative? Does it contribute to knowledge, criticism, or commentary? Commercial use generally weighs against fair use.
2. The nature of the copyrighted work: Is the work creative or factual? Published or unpublished? Creative works generally receive greater copyright protection.
3. The amount and substantiality of the portion used: Is the amount of copyrighted material used necessary for the purpose? Is it a significant portion of the original work?
4. The effect of the use upon the potential market for or value of the protected work: Does the use harm the market for the original work or substitute for it?

Courts will weigh these factors on a case-by-case basis to determine whether the use of copyrighted material for AI training constitutes fair use.

## Emerging Legal Landscape and Case Studies

There is a dearth of legal precedent regarding fair use and AI training. However, a few recent cases offer a glimpse into how courts might approach this issue.

In 2023, a lawsuit was filed by Thomson Reuters against Ross Intelligence<sup>2</sup>, a company developing AI-powered legal research tools. Thomson Reuters argued that Ross Intelligence

---

<sup>1</sup> *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

<sup>2</sup> *Thomson Reuters v. Ross Intelligence* (2023)

infringed upon their copyrights by using legal documents in their training data. The outcome of this case, currently scheduled for trial in 2024, could set a significant precedent for fair use in AI training.

Another case to consider is *Google LLC v. Oracle America, Inc.* (2014)<sup>3</sup>. Here, the Supreme Court ruled that Google's use of a portion of Java SE application programming interfaces (APIs) in their Android operating system constituted fair use. This case is significant because it highlights the transformative nature of using copyrighted material to create a new and functionally distinct work.

### **Potential Solutions and the Future of Fair Use in AI Training**

The ongoing debate surrounding fair use and AI training highlights the need for potential solutions that balance innovation in the AI field with the protection of intellectual property rights. Here are some possibilities to consider:

**Clearer Guidelines and Best Practices:** Developing clear and consistent legal guidelines specifically addressing fair use and AI training can offer much-needed clarity for both developers and copyright holders. These guidelines could outline the types of data considered fair use for training, the permissible amount of copyrighted material, and the importance of transparency in training processes. Additionally, encouraging best practices within the AI development community, such as anonymizing training data or seeking licensing agreements when dealing with significant amounts of copyrighted works, could be valuable.

**Standardization and Data Sharing Platforms:** Standardizing data formats and creating open-source datasets for AI training could reduce the reliance on copyrighted materials. This approach encourages collaboration and reduces the need for individual developers to scrape or collect copyrighted data. Additionally, fostering data sharing platforms where creators can opt-in to contribute their works to specific AI training purposes could provide a controlled environment for innovation while respecting creator rights.

**Licensing Models and Copyright Collectives:** Establishing licensing models specifically tailored for AI training could offer a more structured solution. These licenses could grant developers access to copyrighted data for training purposes while providing fair compensation to copyright holders. Additionally, the creation of copyright collectives representing various creative industries could simplify the licensing process for developers who need access to diverse training data.

---

<sup>3</sup> *Google LLC v. Oracle America, Inc.*, 572 U.S. \_\_\_\_ (2014)

**Legislative Reform:** In some cases, legislative reform might be necessary to address the specific challenges presented by AI training. This could involve revising existing copyright laws to explicitly address fair use in the digital age or creating a new sui generis (unique) right for training data that balances innovation with creator rights.

**Technological Solutions:** Advancements in technology could also play a role in resolving the fair use debate. Techniques for anonymizing training data or obfuscating copyrighted elements within the training process could offer a way to protect intellectual property while allowing for innovative AI development. Additionally, the development of fair use detection algorithms could help identify potential copyright infringement during the training process.

### Finding the Right Balance

The ideal solution likely lies in a combination of these approaches. Fostering open dialogue between AI developers, copyright holders, and policymakers is crucial to ensure a legal framework that promotes innovation while safeguarding intellectual property rights. Ultimately, the future of fair use in AI training hinges on finding a balance that allows both AI technology and creativity to flourish

### References:

1. World Intellectual Property Organization (WIPO), "Copyright," <https://www.wipo.int/copyright/en/>.
2. U.S. Copyright Office, Fair Use <https://www.copyright.gov/>.
3. Fair Use: Training Generative AI, by Stephen Wolfson (2023) (<https://creativecommons.org/2023/02/17/fair-use-training-generative-ai/>)
4. The Future of Fair Use in an AI-Powered World, by Pamela Samuelson (2022) (<https://law.stanford.edu/stanford-lawyer/articles/artificial-intelligence-and-the-law/>)