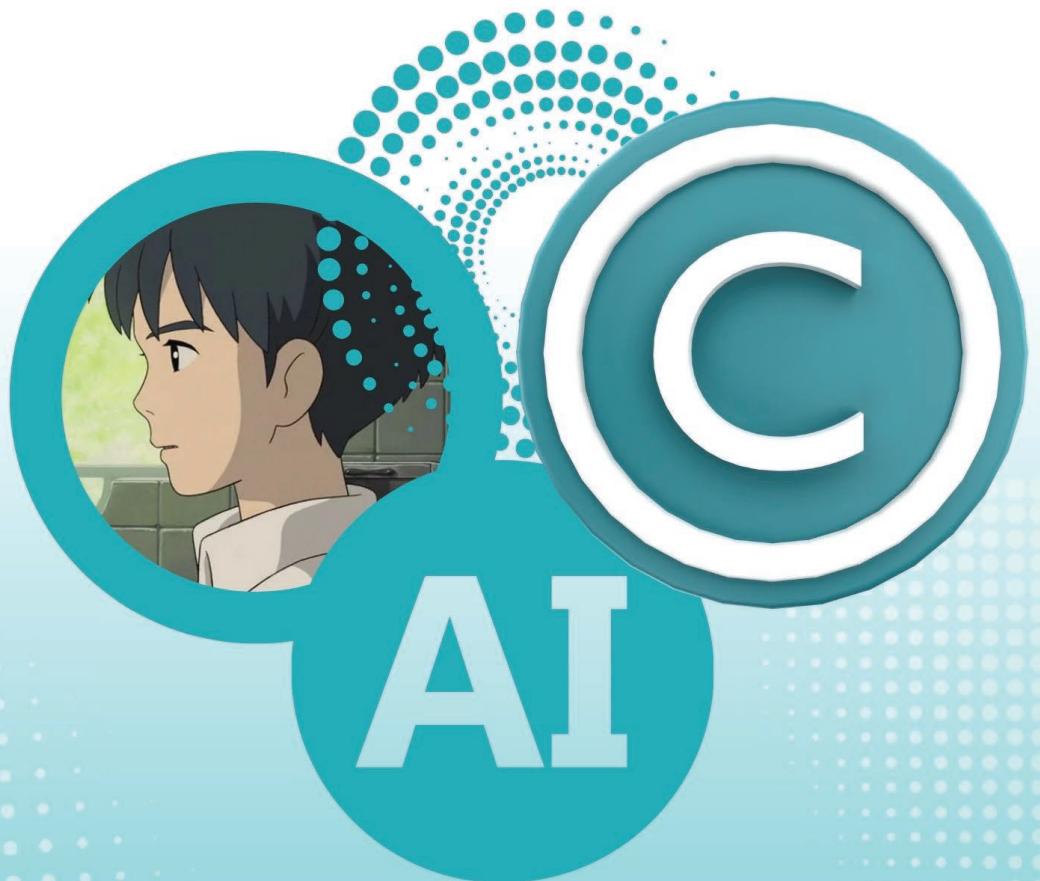




**CoRE-AI**<sup>TM</sup>  
Coalition for Responsible Evolution of AI

Discussion Paper

# Exploring the Intersection of GenAI and Copyright Law



May, 2025

DISCUSSION PAPER

# EXPLORING THE INTERSECTION OF GENAI AND COPYRIGHT LAW

Authors: Jameela Sahiba, Sneha Jain, Akshat Agrawal and Mishthi Dubey

Copyeditor: Akriti Jayant

Thematic designer: Shivam Kulshrestha

Established in July 2024, **CoRE-AI™ (Coalition for Responsible Evolution of AI)** is a prominent multi-stakeholder initiative hosted by The Dialogue, focused on fostering the responsible and ethical development of AI technologies. By bringing together stakeholders from industry, academia, startups, and civil society, CoRE-AI aims to drive collaborative efforts that address the risks associated with AI while maximizing its societal benefits. The initiative seeks to guide India's AI journey, ensuring that technological advancements align with ethical standards to benefit the broader public.

**The Dialogue®** is a public policy think tank with a vision to drive a progressive narrative in India's policy discourse. Founded in 2017, we believe in facilitating well-researched policy debates at various levels to help develop a more informed citizenry, on areas around technology and development issues. The Dialogue® has been ranked as the world's Top 10 think tanks to watch out for, by the Think Tank and Civil Societies Programme (TTCSP), University of Pennsylvania in their 2020 and 2021 rankings.

**For more information**

[www.core-ai.in](http://www.core-ai.in) and [www.thedialogue.co](http://www.thedialogue.co)

**Suggested Citation**

Sahiba, J., Jain, S., Agrawal, A. & Dubey, M. (May 2025). *Discussion Paper: Exploring the Intersection of Gen AI and Copyright Law*. CoRE-AI

**Catalogue No.**

CAI/DP/0525/01

**Publication Date**

May 6, 2025

**Disclaimer**

This discussion paper is an academic exercise conducted to analyse the intersection of Indian copyright act, 1957 with Generative AI technologies

# CONTENTS

<b>1. Introduction to Generative AI</b>	<b>1</b>
1.1 Mode of Learning	1
1.2 Need for Access, Intermediate Copies, and Storage of Expressive Works	2
1.3 Aim and Target of the Technology	2
<b>2. Introduction to Indian Copyright Law and the Key Provisions implicated</b>	<b>3</b>
2.1 Rights of right owners	3
2.2 What Constitutes Copyright Infringement	4
2.3 The Distinction Between Infringement and Fair Dealing	4
2.4 Electronic Storage of Copyrighted Works	5
2.5 Limitations to the Exemptions	5
2.6 Principled Limits on the Scope of Copyright Protection	6
2.7 Liability Framework	6
2.8 Moral Rights	7
<b>3. Intersection of Copyright Law with GenAI Technology</b>	<b>8</b>
3.1 Training stage	8
3.2 Output stage	14
3.3 Liability Frameworks	16
3.4 Moral Rights Framework	17
<b>4. Conclusion</b>	<b>18</b>

# 1. INTRODUCTION TO GENERATIVE AI

Generative Artificial Intelligence (GenAI) marks a transformative leap in computational technology, enabling the creation of diverse forms of content, including text, images, audio, and even video. By harnessing advances in machine learning, especially neural networks, GenAI generates outputs that closely resemble human creativity and cognition.<sup>1</sup> This breakthrough unlocks new levels of creativity and efficiency across industries. However, its rapid adoption also raises pressing legal and ethical concerns, particularly around intellectual property rights. As GenAI continues to evolve, stakeholders grapple with the challenge of balancing technological progress with the need to protect the rights of creators and owners.

GenAI's ability to mimic human creativity depends on vast datasets, many of which include copyright-protected material. Understanding the legal implications of this reliance requires examining how GenAI models function and intersect with intellectual property laws. This discussion paper explores the applicability of various provisions of copyright law to the GenAI landscape and aims to spark dialogue on the need to reimagine the Indian Copyright Act, 1957 (hereinafter referred to as the 'Act') in light of the evolving AI ecosystem.

For the purposes of this discussion paper, our focus is specifically on copyright concerns arising during the training phase of generative AI models. This paper does not delve into the copyright implications at the output stage or the question of ownership in AI-generated works. By narrowing the scope to the training stage, this paper seeks to contribute to a more nuanced understanding of how the Act interacts with the foundational processes that underpin generative AI. This paper is part of a broader series that will examine output-stage questions in subsequent

installments, offering a comprehensive analysis of the intersection between AI and copyright law across different stages.

## 1.1 MODE OF LEARNING

At its core, GenAI operates through a sophisticated learning process that extracts patterns and meta-information from vast datasets, commonly referred to as "training sets." The process involves several stages:

- **Tokenisation and Categorisation<sup>2</sup>:** Data from training sets is deconstructed into fundamental units called "tokens." These tokens can represent words, pixels, or other minimal units of content.
- **Statistical Pattern Recognition<sup>3</sup>:** GenAI algorithms identify statistical patterns among these tokens, analysing their placement and context within the dataset. This stage enables the model to understand relevance, context, and meaning within the data.
- **Prediction and Output Generation<sup>4</sup>:** By applying the patterns and knowledge extracted during training, the model generates outputs that align with the learned context and relevance.

The extracted "knowledge" is a distilled representation of the information embedded within the training sets, rather than a verbatim reproduction of the input data<sup>5</sup>. This essentially means that GenAI produces outputs through what it learns from this distilled knowledge in the training data, mostly avoiding any direct replication of copyrighted works. However, the boundaries between inspiration, adaptation, and reproduction remain nuanced, particularly when GenAI outputs closely resemble existing works.

<sup>1</sup> AnalytixLabs, *Generative AI vs Traditional AI*, May 8th, 2024,

<https://medium.com/@byanalytixlabs/generative-ai-vs-traditional-ai-understand-key-differences-ca2d3e37c45d>

<sup>2</sup> Rajaraman, N., Jiao, J., & Ramchandran, K. (2024). *Toward a theory of tokenization in LLMs*. arXiv. <https://arxiv.org/abs/2404.08335>

<sup>3</sup> Bansal, Vasu, *Pattern Recognition in Artificial Intelligence*, 26 Feb 2024,

<https://www.scaler.com/topics/pattern-recognition/>

<sup>4</sup> Bakir, G., Taskar, B., Hofmann, T., Schölkopf, B., Smola, A., & Vishwanathan, S. V. N. (2007). *Predicting structured data*. Cambridge, MA: MIT Press.

<https://mitpress.mit.edu/9780262528047/predicting-structured-data/>

<sup>5</sup> Art Neill, James Thomas and Erika Lee (Vol. 32:225) *A Framework for Applying Copyright Law to the Training of Textual Generative Artificial Intelligence*

[https://tipj.org/wp-content/uploads/2024/08/32TIPLJ3\\_NeillThomasLee.pdf](https://tipj.org/wp-content/uploads/2024/08/32TIPLJ3_NeillThomasLee.pdf)

## 1.2 NEED FOR ACCESS, INTERMEDIATE COPIES, AND STORAGE OF EXPRESSIVE WORKS

The current methodology for training GenAI models necessitates making intermediate copies of datasets, which may include copyrighted works, during various stages of development. These stages include data collection, preprocessing, and model training.

Under the Act, the exclusive right to reproduction<sup>6</sup> is a cornerstone of copyright protection. Sections 14(a)(i)<sup>7</sup> and 14(c)(i)<sup>8</sup> explicitly confer this right to owners of literary, dramatic, musical, and artistic works, encompassing the storage of such works in electronic forms. Consequently, GenAI's reliance on copying and storing data for training raises significant legal questions.

While the act of copying during training might appear to infringe reproduction rights, it is important to note that GenAI does not aim to reproduce or distribute expressive works as they are. Instead, it seeks to gauge and learn patterns, syntax and meta-information across a vast set of works, and generate outputs based on these patterns. This distinction is critical in evaluating whether the intermediate copying constitutes infringement, particularly if the copies are not directly used for commercial dissemination.

## 1.3 AIM AND TARGET OF THE TECHNOLOGY

Gen AI as a technology is not designed to replicate the expressive elements of its training data<sup>9</sup>. Instead, its purpose is to learn by identifying patterns within the accessed datasets and create outputs that are informed by this learning. The underlying principle rests on the idea that copyright protection extends to the original form of expression of ideas, not the ideas or embedded information and functional triggers themselves.

The process of knowledge extraction in GenAI underscores this distinction. By analysing and internalising statistical patterns, the model generates outputs that reflect its learned understanding rather than directly reproducing input data. This conceptual framework positions GenAI as a transformative tool, emphasising its role in innovation rather than replication.

<sup>6</sup> The right of reproduction in the Copyright Act of 1957 states that no one can make copies of a work or a substantial part of it without the copyright owner's permission. This includes making copies in any material form, such as sound or film recordings.

<sup>7</sup> Government of India. (1957). *The Copyright Act, 1957 Section 14(a)(i)*. <https://indiankanoon.org/doc/1129646/>

<sup>8</sup> Government of India. (1957). *The Copyright Act, 1957 Section 14(c)(i)*. <https://indiankanoon.org/doc/1129646/>

<sup>9</sup> Lin and Partner (2024) How Generative AI is already Challenging Copyright Law: Part 2 <https://www.lexology.com/library/detail.aspx?g=8523f486-5583-48c8-a988-2fc5b308d177>

## 2. INTRODUCTION TO INDIAN COPYRIGHT LAW AND THE KEY PROVISIONS IMPLICATED

### 2.1 RIGHTS OF RIGHT OWNERS

The Act<sup>10</sup> grants copyright holders a range of exclusive rights, which provides a strong foundation for the protection of intellectual property. While these rights are broadly classified, these differ significantly across categories such as literary, dramatic, musical, artistic works, cinematograph films, sound recordings, and computer programs. These rights ensure that authors and copyright holders retain control over the use and distribution of their works. By doing so, the Act seeks to strike a balance between incentivising creativity and serving the broader public interest.

- **Rights in Literary Works**

For literary works,<sup>11</sup> including written texts, books, and computer programs, the copyright owner holds a broad set of exclusive rights. These include the rights to reproduce the work, issue copies to the public, perform the work publicly, and communicate it to the public. Additionally, the owner retains the rights to create derivative works, such as adaptations and translations.

- **Rights in Dramatic and Artistic Works**

Dramatic works,<sup>12</sup> such as plays, scripts, and choreographic compositions, and artistic works, including paintings, drawings, and sculptures, share several key rights. Copyright owners may reproduce these works, communicate them to the public, issue copies, and adapt or translate them. Adaptation rights are specially important, as they allow the transformation of a work from one form to another, for instance converting a literary work into a dramatic script or vice versa.

- **Rights in Musical Works**

Musical works,<sup>13</sup> defined as compositions consisting solely of music, excluding lyrics or spoken words, carry rights that reflect their performative and distributive nature. Copyright

holders have the exclusive rights to reproduce the work, perform it publicly, communicate it to the public, and create sound recordings or cinematograph films based on it. The law also grants rights to adapt and translate musical works, acknowledging the fluid and evolving nature of music as it is reinterpreted across genres, languages, and cultural contexts.

- **Rights in Cinematograph Films and Sound Recordings**

Cinematograph films and sound recordings are unique as composite works involving multiple contributors. Copyright owners hold exclusive rights to reproduce these works, sell or rent copies, and communicate these works to the public. Notably, the law offers granular protection by recognising individual frames of a film as photographs. Similarly, it protects sound recordings from unauthorised copying, distribution, and public communication, highlighting the significant economic and cultural value of these media.

#### 2.1.1 The Right of Reproduction and Communication

At the core of copyright protection lies the right of reproduction,<sup>14</sup> which prohibits unauthorised copying of a work in any material form, including digital formats such as computer memory storage. This right allows copyright owners to control how their work is disseminated and monetised. Complementing this is the right of communication to the public, which grants owners the ability to make their works accessible through broadcasts, streaming, or other forms of public display.

<sup>10</sup>. Government of India. (1957). *The Copyright Act, 1957*. <https://copyright.gov.in/documents/copyrightrules1957.pdf>

<sup>11</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 2(o)*

<sup>12</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 2(h)*

<sup>13</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 2(p)*

<sup>14</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 14(a)(i)* <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>



## 2.1.2 The Concept of Adaptation

The right to adaptation empowers copyright owners to permit or prevent the creation of derivative works that transform or reinterpret the original. These adaptations can take many forms, such as converting a dramatic work into a novel, transcribing a musical composition, or visualising a literary work through visual media such as comics or films. This right safeguards the integrity of original work while ensuring that creators benefit from its evolving expressions and formats.

## 2.1.3 Moral Rights

Beyond the economic rights outlined above, authors retain moral rights,<sup>15</sup> which exist independently of copyright ownership. These rights allow authors to assert authorship and to object to any distortion, mutilation, or modification of their works that could harm their honor or reputation. The Act grants these rights perpetually, ensuring that an author's legacy remains protected even after transferring or assigning their economic rights.

## 2.2 WHAT CONSTITUTES COPYRIGHT INFRINGEMENT

Copyright infringement occurs when the aforementioned rights are violated through unauthorised and prohibited use of protected works. It is critical to recognise that copyright law protects the expression of ideas, not the ideas themselves, emphasising the value of tangible creative output. This distinction preserves the balance between protecting creativity and promoting the free flow of knowledge and innovation. Moreover, not every use of a copyrighted work amounts to infringement; only those that violate the specific rights listed under Section 14 of the Act qualify as infringing acts.

### 2.2.1 Acts Constituting Copyright Infringement

- Copyright infringement encompasses a variety of unauthorised activities that directly or indirectly harm the copyright owner's exclusive rights. Common forms of infringement<sup>16</sup> include:

- **Reproduction for Sale or Hire**

Unauthorised reproduction of a copyrighted work for commercial purposes, such as creating counterfeit books, films, or music recordings for sale or rental, constitutes a blatant infringement. Such acts undermine the economic interests of copyright owners by diverting revenue streams and saturating the market with unlicensed copies.

- **Unauthorised Public Performance**

Permitting a venue for public performance of a copyrighted work, such as a dramatic play or musical composition, without obtaining the necessary licenses from the copyright owner is an infringement. This applies to both physical spaces and digital platforms that facilitate unauthorised performances.

- **Distribution of Infringing Copies**

The act of distributing unauthorised copies of a copyrighted work for trade or in a manner that prejudicially affects the interests of the copyright owner constitutes infringement. This includes both physical distribution, such as pirated DVDs, and digital dissemination through file-sharing platforms.

- **Public Exhibition for Trade**

The public exhibition of infringing copies, such as displaying counterfeit art in galleries or unauthorised screenings of films, is a clear violation of copyright law. Such acts not only harm the copyright owner financially but also compromise the integrity of their work.

- **Importation of Infringing Copies**

Importing unauthorised copies of copyrighted works into India without the consent of the copyright owner is another form of infringement. This provision aims to curb cross-border piracy and protect the domestic market for legitimate copyright owners.

## 2.3 THE DISTINCTION BETWEEN INFRINGEMENT AND FAIR DEALING

Indian copyright law provides exceptions to infringement under Section 52 of the Act, including under the doctrine of "fair dealing," as provided under

<sup>15</sup> Jain, S. (2020). *Moral rights of the author under copyright laws*. SSRN. <https://ssrn.com/abstract=3902409>

<sup>16</sup> Copyright Office. (n.d.). *Handbook of Copyright Law*. Government of India. <https://copyright.gov.in/documents/handbook.html>



Section 52(1)(a) of the Act,<sup>17</sup> thereby allowing limited use of copyrighted works for purposes such as research, criticism, or education. However, these exceptions are narrowly defined and do not extend to commercial exploitation or substantial reproduction. The law's objective is to strike a balance between protecting the rights of creators and enabling access to knowledge and culture. It is important to note here that Section 52 becomes relevant only when an act appears to infringe upon the exclusive rights granted under Section 14 of the Act.

### 2.3.1 Statutory exemptions to copyright infringement

A number of statutory exemptions<sup>18</sup> to the general rule of copyright protection are carved out by the Act under Section 52. By striking a balance between the rights of copyright holders and the broader public interest, these exclusions make sure that the law doesn't unnecessarily restrict freedom of expression, creativity, or access to information. Section 52 lists some situations in which using copyrighted content does not amount to infringement, permitting limited and purposeful engagement with protected works without the owner's consent.

### 2.3.2 Fair Dealing: A Cornerstone of Statutory Exemptions

The principle of fair dealing, as codified under Section 52(1)(a), is a foundational exemption in Indian copyright law. It allows limited use of copyrighted works for purposes deemed to serve the greater public good. Unlike the broader "fair use" doctrine under U.S. law, fair dealing under Indian law is more restrictive, specifically listing permitted uses. These include:

- **Private or Personal Use, Including Research**

Individuals may engage with copyrighted works for private study or personal use without constituting infringement. This exemption is particularly significant in educational and academic contexts, allowing students, researchers, and scholars to access and analyse works without requiring explicit permission from copyright holders. It also facilitates the storage of such works in electronic formats for personal use, provided they are not reproduced or distributed for commercial purposes.

- **Criticism or Review**

Fair dealing permits the use of copyrighted material for critical or analytical purposes, such as book reviews, art critiques, or scholarly evaluations. This provision underscores the importance of fostering intellectual discourse and ensuring that copyright does not obstruct legitimate critique or scholarly debate. These kinds of uses frequently only include a few quotes or references that don't significantly replicate the original work.

- **Reporting of Current Events and Affairs**

Copyright law also accommodates the reporting of current events and public lectures, recognising the societal value of journalism and public discourse. For instance, reproducing excerpts from speeches or quoting sections of literary works for news reporting does not amount to infringement, as long as it is fair and proportionate to the purpose of reporting.

## 2.4 ELECTRONIC STORAGE OF COPYRIGHTED WORKS

An important addition to the statutory exemptions is the explicit recognition of electronic storage of works for permitted purposes. The explanation to Section 52(1)(a) clarifies that storing copyrighted material in electronic form, such as digital archives or personal electronic devices, for private use, research, or criticism does not infringe copyright. This acknowledgment reflects the evolving nature of content consumption in the digital era and aligns copyright law with technological advancements.

## 2.5 LIMITATIONS TO THE EXEMPTIONS

While Section 52 provides crucial safeguards for public access and creative engagement, these exemptions are not absolute. All statutory exemptions are based on the fairness principle, which makes sure that using content protected by copyright doesn't interfere with the work's normal utilisation or infringe upon the legitimate interests of the copyright owner. For example, using an entire book for personal research may exceed the bounds of fair dealing, especially if there is significant reproduction that lowers the work's market worth.

17. Ahmed, S. (2021). *Fair Dealing in Indian Copyright Law*. *Journal of Intellectual Property Rights*, 26(2), 96–102.

<https://nopr.niscpr.res.in/handle/123456789/58020>

18. Government of India. (1957). *The Copyright Act, 1957 Section 52*. <https://copyright.gov.in/documents/copyrightrules1957.pdf>

## 2.6 PRINCIPLED LIMITS ON THE SCOPE OF COPYRIGHT PROTECTION

The Act serves to protect the rights of creators over their original works while balancing these rights with broader societal and legal principles. One of the key principles underlying copyright law is its limitation to expressions of ideas, rather than the ideas themselves. This distinction, often referred to as the idea-expression dichotomy, serves as a cornerstone of copyright law and establishes principled boundaries on the scope of copyright protection.

### 2.6.1 The Idea-Expression Distinction: Delineating the Scope of Copyright

Indian copyright protection rests on the requirement of "originality." Although the Act does not explicitly define the term "original," judicial interpretation has shaped its meaning. Originality means that the work must "owe its origin to the author," indicating that it results from the author's own skill, judgment, and labor, rather than being a mere replication of existing works. Another essential criterion for copyright protection is fixation of the work in a tangible medium, which ensures that the expression of the work is concretely manifested.

India's copyright framework follows the foundational principles of international copyright law, as enshrined in Article 9(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)<sup>19</sup> and Article 2 of the WIPO Copyright Treaty (WCT),<sup>20</sup> 1996. These provisions affirm that copyright protects only the original expression of those ideas, not the ideas, concepts, or abstract notions themselves. This distinction ensures that copyright law incentivises creativity without hindering the free exchange and development of foundational ideas. For instance, while the narrative style or character development in a novel qualifies for protection, the broader themes or underlying ideas remain in the public domain. This balance fosters innovation and creativity without monopolising the essential building blocks of knowledge or artistic expression. By safeguarding only the tangible form of expression, copyright law

encourages others to draw inspiration from and build upon existing ideas in new and original ways. That said, applying the idea-expression concept to the AI domain may also necessitate a more careful analysis, particularly in scientific, academic, or computer program works,<sup>21</sup> wherein sometimes the expression of the idea may closely resemble the idea itself.

### 2.6.2 Economic Rationale for Copyright: Protecting the Primary Market

Copyright law primarily seeks to prevent unauthorised reproduction that could replace or compete with the copyright owner's original market. This economic underpinning ensures that creators receive fair recompense for their intellectual contributions. However, this right is not absolute and remains subject to doctrinal limitations. Courts have held that copyright protection does not cover activities that do not impact the primary market or the copyright holder's economic returns.<sup>22</sup> For example, merely storing or copying a work without exposing its expressive elements to public consumption may not constitute infringement,<sup>23</sup> as it does not attempt to undermine the work's market value.

## 2.7 LIABILITY FRAMEWORK

The Act establishes out a comprehensive liability framework to protect the rights of copyright owners and deter infringement. It includes both civil and criminal liabilities, offering remedies and sanctions against unauthorised use of copyrighted works. The system ensures accountability while balancing the interests of creators, users, and the broader public.

### 2.7.1 Civil Liability for Copyright Infringement

The Act empowers copyright owners to initiate civil actions<sup>24</sup> against individuals or entities that infringe their rights. Common forms of infringement include making or distributing unauthorised copies, performing copyrighted works publicly without permission, and importing infringing copies into India. Civil remedies

<sup>19</sup> World Trade Organization. (1994). *Agreement on Trade-Related Aspects of Intellectual Property Rights*. <https://www.wipo.int/wipolex/en/treaties/details/231>

<sup>20</sup> World Intellectual Property Organization. (1996). *WIPO Copyright Treaty (WCT)*. [https://www.wipo.int/edocs/pubdocs/en/wipo\\_pub\\_226.pdf](https://www.wipo.int/edocs/pubdocs/en/wipo_pub_226.pdf)

<sup>21</sup> Harsha Pathak and Anindya Sircar *Idea-Expression Dichotomy: Challenges to Computer Programs* <https://ijalr.in/volume-3-issue-4/idea-expression-dichotomy-challenges-to-computer-programs-harsha-pathak-anindya-sircar/>

<sup>22</sup> The Chancellor Masters & Scholars Of The ... vs Narendera Publishing House And Ors, on 17 September, 2008 <https://indiankanoon.org/doc/13819251/>

<sup>23</sup> R.G Anand vs M/S. Delux Films & Ors on 18 August, 1978 <https://indiankanoon.org/doc/1734007/>

<sup>24</sup> e-PG Pathshala. (n.d.). *Intellectual Property Copyright- Civil and Criminal Remedies*.

[https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp\\_content/law/08.\\_intellectual\\_property\\_law/40.\\_copyright\\_civil\\_and\\_criminal\\_remedies/et/8099\\_et\\_et.pdf](https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/law/08._intellectual_property_law/40._copyright_civil_and_criminal_remedies/et/8099_et_et.pdf)

available to copyright owners include injunctions to prevent further infringement, damages for economic and reputational harm, and the disgorgement of profits earned through infringement.

### 2.7.2 Criminal Liability for Copyright Infringement

The Act also recognises copyright infringement as a criminal offence under Section 63<sup>25</sup>. Any individual who knowingly infringes or abets infringement faces criminal prosecution. The threshold for criminal liability is intent or knowledge of infringement. Criminal acts include unauthorised reproduction or distribution of copyrighted works, circumvention of technological measures, removal of rights management information, and false representations to deceive authorities.

## 2.8 MORAL RIGHTS

Moral rights form an essential component of copyright law in India, preserving an author's personal connection with their work regardless of economic rights. Section 57<sup>26</sup> of the Act grants these rights, which remain independent of economic interests and continue to exist even after the copyright has been assigned, sold, or expired.

---

<sup>25</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 63*. <https://copyright.gov.in/documents/copyrightrules1957.pdf>

<sup>26</sup>. Government of India. (1957). *The Copyright Act, 1957 Section 57*. <https://copyright.gov.in/documents/copyrightrules1957.pdf>

# 3. INTERSECTION OF COPYRIGHT LAW WITH GENAI TECHNOLOGY

Copyright law has traditionally aimed to protect creators' rights while encouraging innovation. However, the rise of GenAI presents new complexities that test the boundaries of existing legal frameworks. This chapter explores these intersections, examining how GenAI's dependence on expressive works for training and its ability to generate similar content create legal grey areas. It unpacks the tension between protecting creators' intellectual investments and promoting technological advancement, while analysing copyright law's applicability to the GenAI landscape.

## 3.1 TRAINING STAGE

The training stage of GenAI models relies fundamentally on datasets curated after accessing and scraping vast amounts of content and information available over the internet. This initial process of accessing and scraping content forms the crucial foundation of the training dataset for the model, raising important legal and ethical considerations within the framework of copyright law.<sup>27</sup>

### 3.1.1 Access

- **Access to Publicly Available Copyrighted Works**  
"Publicly available copyrighted works" refers to expressions that are accessible without bypassing access restrictions or rights management systems, and without needing the copyright owner's consent. For instance, a film uploaded to YouTube by its rightful owner remains protected despite being publicly accessible. These must be distinguished from works in the 'public domain,' which is the term used to describe creative works, such as music, photographs, or poems, that are no

longer covered by copyright because of expiration, forfeiture, or inapplicability. This implies that anyone can utilise works in the public domain for any use without requesting permission.

This distinction becomes particularly relevant when comparing personal use to automated access. When a person buys a book for private enjoyment, they engage with the content in a fundamentally different way than a GenAI model scraping publicly available content for training purposes.<sup>28</sup> While copyright law protects the specific expression of a work, it does not extend to the underlying ideas. Since publicly available works are intentionally shared by the copyright owner, accessing them does not constitute trespass<sup>29</sup>, nor does it diminish the copyright holder's control over their market. Unlike the taking of physical property, digital access to publicly available works does not diminish the copyright holder's protected market of their work.

Moreover, GenAI does not "read" or derive meaning in the conventional sense. Instead, it tokenises content, transforming scraped statements into numerical representations that the model can understand.<sup>30</sup> The AI creates a logical continuation of the input by examining patterns in these tokens and mapping them onto the format of user prompts. Given the vast volume of training data, the likelihood of the model reproducing any single tokenised work in a way that amounts to meaningful resemblance is extremely low.

- **Technological Protection Measures (TPMs) and Legal Framework**

Although many GenAI models rely on publicly available data, often supplemented by

<sup>27</sup> Sneha Jain and Akshat Agrawal. Indian Copyright Law and Generative AI. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai/>

<sup>28</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI-Part 5: Right to exclude access?*. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-5-right-to-exclude-access/>

<sup>29</sup> The "trespass argument" refers to the analogy between accessing copyrighted content in the digital realm and physical trespass onto someone's property. In this context, the argument posits that accessing copyrighted works without explicit permission—whether for training generative AI models or other purposes—could be likened to trespassing on private property, where unauthorized entry violates the owner's exclusive right to control access. However, in the context of AI, the "trespass argument" often falters because the digital access of publicly available works for purposes like training models generally focuses on extracting unprotected elements (e.g., ideas or facts) rather than infringing on the protected expression. Thus, while the trespass analogy might resonate on an intuitive level, it fails to account for the nuanced framework of copyright law, which allows for lawful access and use in specific contexts.

<sup>30</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI-Part 5: Right to exclude access?*. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-5-right-to-exclude-access/>

proprietary<sup>31</sup> conversational datasets held by companies, Technological Protection Measures (TPMs), such as paywalls or encryption, play a significant role in this context. Right holders use these measures to restrict access to copyrighted works, creating both legal and technological safeguards against unauthorised uses.

Section 65A<sup>32</sup> of the Act, 1957, directly addresses the circumvention of TPMs, explicitly prohibiting the act of bypassing 'effective' protection mechanisms designed to safeguard the rights conferred by copyright. This provision reflects India's alignment with the World Intellectual Property Organisation (WIPO) Copyright Treaty and the global push to strengthen digital rights management. However, the provision's scope is not without limitations. It prohibits circumvention only when done with the intention of infringing copyright, thereby introducing a critical *mens rea* (guilty mind) element to the analysis.<sup>33</sup> For instance, bypassing a paywall with the specific intention to reproduce or distribute a copyrighted work clearly falls within the scope of this prohibition. In contrast, the mere act of circumvention without the intent to infringe does not amount to violation.

The scope of this paper, however, is restricted to the access and use of publicly available copyrighted works used in the process of training. It does not cover the use of TPM-protected content accessed by breaking such TPMs or by retrieving works from illegal websites or pirated sources. This paper does not endorse the use of paywalled data obtained through illegally circumventing paywalls or from other unauthorised and pirated sources.

#### • Private Use Defence and Commercial Considerations

The private use defence, which is provided by Section 52(1)(a)(i) of the Act, is widely understood to include non-commercial usage that does not entail dissemination/publication for exposure.<sup>34</sup> "Fairness" in dealing with work for private use

requires a transformed purpose coupled with non-market substitutability. In the context of GenAI, since developers do not directly use the extracted data in its original form for profit, using copyrighted content in training may be considered non-commercial. Rather than duplicating or disseminating the copyrighted expression itself, the training process concentrates on identifying underlying patterns, concepts, or factual elements, elements that are not protected by copyright.

This approach resembles a lawyer privately using publicly available legal commentary, where the initial use is non-commercial, even if the knowledge later benefits the lawyer in their career. A similar case arises when a journalist consults copyrighted material, such as archived news or historical commentaries, to gather background information for an investigative story. While the final report may carry commercial value, the initial access remains transformative, non-commercial, and essential for informed analysis.

Having said that, it is important to recognise that the legal interpretation of "private use" in GenAI contexts is not a settled legal position and necessitates a thorough analysis of the particular facts and circumstances of each case. Thus far, India's attitude seems especially lenient when it comes to web-scraping for AI training. It is arguable that web scraping may not fall under the conventional purview of copyright protection when it is used to retrieve information or learn from a work without reproducing its expressive form. This claim, however, depends on the kind of content accessed, the nature of the scraping activity, and whether the procedure infringes upon the exclusive rights of the copyright owner as laid out in the Act.

This approach offers greater flexibility for AI and machine learning in the digital age by prioritising purpose and intent over mere access. Even with paywalls in place, it may be difficult to hold training dataset compilation by web scraping liable because the emphasis is on preventing piracy

<sup>31</sup> Data essentials in the age of GenAI(2024) <https://www.accenture.com/hu-en/insights/data-ai/new-data-essentials>

<sup>32</sup> Government of India. (1957). *The Copyright Act, 1957 Section 65A*, <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>33</sup> Indian Copyright Law and Generative AI- Part 5: Right to exclude access?(2024) <https://www.lexology.com/library/detail.aspx?g=2515f6a9-2944-4893-a9ec-be9a0157d6ff>

<sup>34</sup> Government of India. (1957). *The Copyright Act, 1957 Section 52(1)(a)(i)*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>



rather than legal extraction. This is valid provided that the scraping does not violate any rights guaranteed by Section 14<sup>35</sup> of the Act.

On the contrary, it is also important to note here that the right to refuse, which is essentially the ability for content publishers/owners to refuse the scraping of their data for purposes of AI training, is also being adopted as a means to protect copyrightable data. The same has been recognised as the right to 'opt-out' of the text and data mining (TDM) exception under Article 4 of the Copyright in the Digital Single Market (CDSMD) Directive<sup>36</sup> in the EU. In India, while there is no binding directive on the same, some websites owners are using robots.txt files to deter AI crawlers from accessing and using their content for training purposes.

### 3.1.2 Intermediate Copying and storage for the purposes of training

- **Understanding the training process**

Training generative AI models involves an essential intermediary step in which developers copy and store datasets, including copyright-protected works, in a process known as 'fixing' during AI development. These datasets are typically stored on centralised cloud servers. This storage may occur in one of three distinct ways: continuously throughout the model's lifecycle, temporarily until the model processes the data, or via federated learning systems that bypass centralised storage altogether.

A critical point in this process is that these developers use these intermediate copies solely to extract meta-information from the content's expression. The original works are not exposed to users or made publicly available in any form during training.

- **The Distinction Between Copying and Storage**

Under copyright law, copying and storing are treated as distinct acts involving the use of protected work.<sup>37</sup> When training generative AI

models, the system reads content to tokenise it and weigh model parameters, determining the logic of potential sequence patterns. This process fundamentally differs from the traditional consumption of copyrighted work - it doesn't involve reading, seeing, hearing, or enjoying the work in its intended primary context. This distinction proves crucial when examining legal implications.

- **Core Rights Under Copyright Law and Limiting doctrines**

Indian copyright law grants exclusive reproduction rights to owners of literary, dramatic, musical, and artistic works, as well as sound recordings and cinematographic films (making a copy right).<sup>38</sup> These rights to control copying also extend to performers' rights and broadcast reproduction rights. Although the Act does not explicitly define "reproduction" or "copy," courts, most notably in *MRF v. Metro Tyres*, have interpreted these terms broadly to cover not just physical duplication but also substantive imitation.<sup>39</sup> This right fundamentally protects copyright owners from losing economic returns due to market substitutes.

However, several court-developed doctrines limit the scope of reproduction rights, and they are specially relevant when considering AI training. The idea-expression distinction, recognised by the Supreme Court in *R.G. Anand v. Delux Films*, establishes that copyright protects only expressive forms, not underlying ideas.<sup>40</sup> Protection covers the unique "packaging" of the idea, and not the idea itself. The Calcutta High Court reaffirmed this principle in *Barbara Taylor Bradford v. Sahara Media*, emphasising that while expression is protected, embedded ideas remain unprotected.<sup>41</sup>

The merger doctrine further limits protection where ideas can only be expressed in limited ways or are functional to the genre, while the de minimis rule excludes protection for trivial usage not constituting a substantial part of the expressive form.

<sup>35</sup> Government of India. (1957). The Copyright Act, 1957 Section 14. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>36</sup> Gina Maria Ziaja (2024) The text and data mining opt-out in Article 4(3) CDSMD <https://academic.oup.com/jiplp/article/19/5/453/7614898>

<sup>37</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI*. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai/>

<sup>38</sup> Government of India. (1957). *The Copyright Act, 1957 Section 14(a)(i)*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>39</sup> Delhi High Court. *MRF v. Metro Tyres*, (2019) SCC Online Del 8973. Also at <https://indiankanoon.org/doc/78110318/>

<sup>40</sup> Supreme Court of India. *R.G. Anand v. Delux Films*, 1978 (4) SCC 118. Also at <https://indiankanoon.org/doc/1734007/>

<sup>41</sup> Calcutta High Court. *Barbara Taylor Bradford and Anr. v. Sahara Media Entertainment Ltd. And Ors.*, 2003 SCC Online Cal 323. Also at <https://indiankanoon.org/doc/757852/>

These doctrines ensure a careful balance between “public interest” in extracting knowledge for the purposes of dissemination and use, and in the commercial enablement interests of those who create original expression encompassing such ideas and knowledge in their own distinct and original way. The Supreme Court of India in *ENIL v. Super Cassettes*<sup>42</sup>, emphasised on this facet of public interest stating:

*“The right to property, therefore, is not dealt with its subject to restrict when a right to property creates a monopoly to which public must have access, withholding the same from public may amount to unfair trade practice. In our constitutional Scheme of statute monopoly is not encouraged. Knowledge must be allowed to be disseminated. An artistic work if made public should be made available subject of course to reasonable terms and grant of reasonable compensation to the public at large.”*

- **Storage Rights Framework**

The 1994 Amendment to the Act explicitly incorporated storage within the scope of reproduction rights to align with India’s obligations under the TRIPS Agreement.<sup>43</sup> Later, the 2010 Parliamentary Standing Committee Report emphasised concerns about unauthorised storage by Internet Service Providers that could lead to impermissible exposure of copyrighted works.<sup>44</sup>

A purposive interpretation of “Reproduction, including storing of works” under Section 14 might exclude storage solely for meta-information extraction (protected by the idea-expression dichotomy), while a literal interpretation could extend the copyright owner’s market to mere storage, regardless of purpose. These are however, in any case, subject to certain limiting principles.

The Act’s framework defines infringement through Sections 14 and 51, establishing that infringement

occurs only when an exclusive right is violated. Section 2(m) further clarifies that an “infringing copy” refers to a reproduction or copy made or imported in contravention of the Act, implying that permitted reproductions fall outside the ambit of exclusive rights and do not constitute “infringing copies.”<sup>45</sup> In *M/s Blackwood and Sons Ltd. v. A.N. Parasuraman*, the Madras High Court interpreted infringing reproduction as requiring publication, thereby distinguishing it from acts of private use.<sup>46</sup> Section 3 of the Act defines publication as “making a work available to the public by issue of copies or by communicating the work to the public,” underscoring that infringement concerns the existence of another work made available to the public, rather than mere internal use of the original.<sup>47</sup> In the case of AI training, no new work is created or published; rather, the process involves the amalgamation of knowledge, not the production or dissemination of a new copyrighted expression.

- **The Transformative Use Doctrine and Application**

The doctrine of transformative use, a significant limiting principle rooted in the Act, plays a crucial role in defining the scope of copyright protection. While the concept was prominently developed in US copyright law, Indian courts have distinctly recognised its application. In *Syndicate of The Press for the University of Cambridge v. B.D. Bhandari*, the Division Bench of the Delhi High Court made significant observations regarding transformative use as a limitation to copyright protection.<sup>48</sup> The court recognised that when the use of a work of “transformative character” - meaning the purpose served differs from the work’s original purpose or transforms the expressive form of the work in a non-trivial manner—it is beyond the subject matter of copyright protection.<sup>49</sup>

In *B.D. Bhandari*, the Court notably referred to the U.S. decision in *Perfect 10, Inc. v. Amazon.com, Inc.*, which defined transformative use as use that does not merely supersede the objectives of the

<sup>42</sup> Supreme Court of India, *M/S. Entertainment Network (India) Ltd vs M/S. Super Cassette Industries Ltd* 2008(37) PTC 353 (SC), Also at <https://indiankanoon.org/doc/1592558/>

<sup>43</sup> The Copyright (Amendment) Act, 1994. [https://copyright.gov.in/Documents/Notification/Copyright\\_Amendment\\_1994.pdf](https://copyright.gov.in/Documents/Notification/Copyright_Amendment_1994.pdf)

<sup>44</sup> Parliament of India. Standing Committee Report on the Copyright Amendment Bill 2010. <https://prsindia.org/billtrack/the-copyright-amendment-bill-2010>

<sup>45</sup> Government of India. (1957). *The Copyright Act, 1957 Section 2(m)*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>46</sup> Madras High Court. *M/s Blackwood and Sons Ltd. And Ors. v. A.N. Parasuraman and Ors.*, 1958 SCC Online Mad 62. Also at <https://indiankanoon.org/doc/1685540/>

<sup>47</sup> Government of India. (1957). *The Copyright Act, 1957 Section 3*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>48</sup> Delhi High Court. *Syndicate of The Press for the University of Cambridge on behalf of the Chancellor, Masters and School v. B.D. Bhandari and Anr.*, 2011 SCC Online Del 3215. Also at <https://indiankanoon.org/doc/565788/>

<sup>49</sup> Delhi High Court. *Syndicate of The Press for the University of Cambridge on behalf of the Chancellor, Masters and School v. B.D. Bhandari and Anr.*, 2011 SCC Online Del 3215. Also at <https://indiankanoon.org/doc/565788/>



original work, but instead adds something new, with a different purpose or character, altering the original with new expression, meaning, or message.<sup>50</sup> The Court's recognition of guidebooks as transformed works, rather than infringing reproductions, provides a crucial precedent for understanding and applying the transformative use doctrine within the Indian copyright framework.

This interpretation finds further reinforcement in the Division Bench judgement of the Calcutta High Court in *Barbara Taylor Bradford v. Sahara Media Entertainment*, where the Court affirmed that when original works are taken and substantially transformed to create entirely different works, with non-trivial variations, such use would not generate actionable claims for the owner.

- **The Transient Storage exception**

Section 52(1)(b) of the Act provides a crucial exemption for transient or incidental storage carried out as part of a technical transmission process.<sup>51</sup> The Delhi High Court, in *Myspace Inc. v. Super Cassettes Industries Ltd.*, interpreted this provision to cover temporary or subordinate storage that merely enhances the core function performance..<sup>52</sup> The exemption could potentially extend to AI training processes where:

- Only cached copies are created incidental to machine learning operations;
- Background copies remain temporary and are solely used to facilitate model training;
- Such copies are never made publicly available; and
- Storage is purely transient or incidental to the model's integration.

- **Implications for AI Training**

In the context of AI training, the above legal principles suggest that copying and storage for non-human computational processing, particularly when transformed for different purposes than the original work, may arguably fall outside traditional copyright restrictions. The emphasis on publication and human exposure within the legal framework, coupled with the recognition of transformative use and technical processing exemptions, creates viable pathways for legitimate AI training activities, while still preserving copyright

### 3.1.3 Technical Implementation of Non-Human Exposure

AI training systems incorporate robust technical measures to ensure that training data remains solely accessible for computational processing, while preventing any human exposure. These safeguards include:

**Security Architecture:**

- **Encrypted Storage:** Data is maintained exclusively in encrypted formats; accessible only to training algorithms
- **Automated Processing:** End-to-end automated pipelines handle data ingestion and processing without human intervention
- **Access Controls:** Multi-layered technical barriers prevent any human access to raw training data

**Technical Safeguards:**

- **One-Way Transformation:** Irreversible tokenisation techniques prevent reconstruction of the original content
- **Secure Enclaves:** Hardware security modules safeguard data during processing
- **Federated Architecture:** Processing occurs on distributed devices without central storage

**Verification Systems:**

- **Automated Logging:** Systems track all data access and processing
- **Access Verification:** Multi-factor authentication and role-based controls
- **Regular Audits:** Independent verification of non-exposure protocols

### 3.1.4 Intermediate Copying to produce competing works

- **The Idea-Expression Distinction**

The Supreme Court of India, in *R.G. Anand v. Deluxe Films*, established a fundamental limitation on copyright owners' exclusive rights through the idea-expression distinction. Under this principle, ideas, principles, themes, subject matters, and historical or legendary facts fall outside the scope of copyright protection. Only the expressive form of a work, and not the embedded knowledge or

<sup>50</sup> United States Court of Appeals for the Ninth Circuit. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146. <https://cdn.ca9.uscourts.gov/datastore/opinions/2007/12/03/0655405.pdf>

<sup>51</sup> Government of India. (1957). *The Copyright Act, 1957 Section 52(1)(b)*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>52</sup> Delhi High Court. *Myspace Inc. v. Super Cassettes Industries Ltd.*, 2016 SCC Online Del 6382. Also at <https://indiankanoon.org/doc/12972852/>

information, is protected. For a claim of copyright infringement to succeed, there must be substantial qualitative or quantitative similarity in expression form, judged from the perspective of a lay observer.

The Court further reinforced this principle by approving the Allahabad High Court's decision in *S.K. Dutt v. The Law Book Company*, which stressed that individuals are free to utilise and build upon the useful elements of an original work.<sup>53</sup> The Court explicitly observed that "*every man can take what is useful from the original work, improve, add and give to the public the whole, comprising the original work, with the additions and improvements.*"

This landmark ruling firmly established that copyright confers primarily a negative right, a right to prevent unauthorised appropriation of one's labour, rather than a monopoly over the embedded knowledge. Multiple individuals may create similar works without infringement, provided they originate independently rather than through blatant copying or colorable imitation of existing works. The burden of proving such unauthorised appropriation lies squarely on the plaintiff.

Any limitation to the idea-expression doctrine in the context of generative AI training, which merely extracts knowledge from training copies, would have significant negative spillover effects across other domains governed by copyright law. It would risk unsettling the foundational balance in public interest which in fact allows conferral of statutory exclusionary rights.

- **Protection of Facts and information**

Indian courts have consistently held that facts embedded within expressive works are not eligible for copyright protection, as they are not products of the author's creativity. The Supreme Court's ruling in *Eastern Book Company v. D.B. Modak* decisively addressed this issue in the context of law report publishing.<sup>54</sup> The case examined claims for protection over publisher's skill and effort in copy-editing judgments through cross-referencing, citation insertion, margin

headings, and typographical corrections. The Court narrowly defined copyright protection, limiting it to elements that meet the modicum of creativity test, which requires demonstrable knowledge, sound judgment, and legal skill, rather than mere effort in copy-editing factual information.<sup>55</sup>

The Delhi High Court further clarified this principle in *Akute Internet Services v. Star India*, reaffirming that copyright protects only expression, not the underlying facts.<sup>56</sup> The court rejected attempts to establish quasi-property rights over public domain information, citing Section 16's preemption doctrine. This decision, which involved sports broadcasting rights over cricket match score updates, demonstrates Indian courts' reluctance to extend protection beyond traditional copyright boundaries, even when faced with "hot-news" doctrine claims.<sup>57</sup>

- **The Extractive Use Defence and Competition**

The extractive use defence addresses the reproduction of publicly available copyright-protected works for three specific exempted purposes: extracting unprotectable ideas, facts, or themes; conducting statistical analysis of tokenised outputs; and producing distinct outputs that lack substantial similarity to the original works. This approach argues that such activities arguably fall outside the scope of copyright protection, as protecting them would impermissibly extend copyright to ideas and facts. The Delhi High Court's decision in *Institute for Inner Studies v. Charlotte Anderson* supports this position, particularly regarding books describing useful arts or historical facts.<sup>58</sup> The court established that while copyright protects specific language and arrangement, it cannot monopolise underlying techniques or knowledge available in the public domain.

This framework actively promotes beneficial competition by ensuring that useful elements or meta knowledge embedded within copyrighted works remain accessible for public use and improvement. When applied to intermediate copying, it suggests that using protected works to

<sup>53</sup> Allahabad High Court. *S.K. Dutt v. The Law Book Company and Ors.*, 1953 SCC Online All 286. Also at <https://indiankanoon.org/doc/146202/>

<sup>54</sup> Supreme Court of India. *Eastern Book Company v. D.B. Modak*, AIR 2008 SC 809. Also at <https://indiankanoon.org/doc/1062099/>

<sup>55</sup> Supreme Court of India. *Eastern Book Company v. D.B. Modak*, AIR 2008 SC 809. Also at <https://indiankanoon.org/doc/1062099/>

<sup>56</sup> Delhi High Court. *Akute Internet Services Pvt. Ltd. & Anr. v. Star India Pvt. Ltd. & Anr.*, 2013 SCC Online Del 3344. Also at <https://indiankanoon.org/doc/66104323/>

<sup>57</sup> Delhi High Court. *Akute Internet Services Pvt. Ltd. & Anr. v. Star India Pvt. Ltd. & Anr.*, 2013 SCC Online Del 3344. Also at <https://indiankanoon.org/doc/66104323/>

<sup>58</sup> Delhi High Court. *Institute for Inner Studies v. Charlotte Anderson*, 2014 SCC Online Del 136, Also at <https://indiankanoon.org/doc/6236790/>

extract ideas, conduct analysis, and create new, distinct outputs may arguably avoid infringement, provided the resulting work does not substantially reproduce the original's protected expression.<sup>59</sup> This freedom to extract unprotected elements fosters innovation by enabling competitors to build upon existing knowledge while developing their own original expressions, ultimately benefiting the public through improved products and services.

This approach aligns with foundational judicial wisdom on knowledge sharing and innovation.<sup>60</sup> The Lahore High Court's prescient 1938 observation that copyright cannot "*close all the avenues of research and scholarship and all frontiers of human knowledge*" remains particularly relevant today.<sup>61</sup> These principles suggest that attempts to restrict AI systems from learning from underlying knowledge within protected expressions may contradict the fundamental purpose of copyright law's idea-expression distinction.

## 3.2 OUTPUT STAGE

### 3.2.1 The Reproduction Right

In analysing generative AI outputs, courts apply the substantial similarity test to determine whether the model's output effectively substitutes the copyright owner's primary market. This becomes particularly relevant when AI systems generate content that might incorporate elements of training dataset expressions, whether as embedded ideas or direct expression. The fundamental question is whether the AI's output creates market substitution for the original works.

- **The Substantial Similarity Framework for AI Outputs**

The Supreme Court of India's test in *R.G. Anand* plays a crucial role in evaluating AI-generated content: would a reader, spectator, or viewer, after experiencing both works, clearly perceive the AI's

output as a copy of the original? The court specifically noted that when material and broad dissimilarities exist, which negate copying intention, and coincidences are clearly incidental, no infringement occurs. For genAI systems, this means that outputs must be evaluated not only for similarities but for whether these similarities arise from legitimate algorithmic learning as opposed to mere reproduction.<sup>62</sup>

The assessment requires both quantitative and qualitative analysis. As established in *Ram Sampath v. Rajesh Roshan*, even small but crucial segments can constitute infringement, particularly when repeatedly used.<sup>63</sup> The Bombay High Court outlined key factors: identifying similarities and differences, determining whether the subsequent work could meaningfully exist without the copied portion, and identifying the "soul" or essential parts of the original work. The court's observation that "*what is worth copying is prima facie worth protecting*" is particularly relevant when AI systems incorporate specific elements from training data.<sup>64</sup>

- **Understanding and Evaluating Regurgitation Risk**

The risk of infringement increases significantly when AI outputs regurgitate expressive forms from training data. Copyright owners must demonstrate both quantitative and qualitative similarities between the AI-generated content and original works. The burden lies in showing that copied elements constitute essential parts, leading to an unmistakable impression of copying. When viewed holistically, similarities must be substantial enough to create an impression of colorable imitation, with differences being merely superficial.<sup>65</sup>

A substantial similarity analysis involves comparing the output produced by the AI model with the input work. Only if the output incorporates or reproduces the primary expressive components of the input work, when compared as a whole, would a finding of substantial similarity that amounts to infringement emerge. Thus, if genAI training

<sup>59</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI: Part 2 – Transformative and Extractive Use*.

[https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-2-transformative-and-extractive-use/#\\_ednref1](https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-2-transformative-and-extractive-use/#_ednref1)

<sup>60</sup> Molly Shaffer Van Houweling, *The Freedom to Extract in Copyright Law*, Journal of Copyright Society (March, 2024) (Unpublished, draft on file with author).

<sup>61</sup> Lahore High Court, Kartar Singh Giani v. Ladha Singh, AIR 1934 Lahore 777; Delhi High Court, Chancellor, Masters and Scholars of the University of Oxford and Ors. v. Rameshwari Photocopy Services and Ors., 235 (2016) DLT 409. Also at <https://indiankanoon.org/doc/114459608/>

<sup>62</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI: Part 3 – The Output Stage: Analyzing Reproduction and Adaptation*. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-3-the-output-stage-analyzing-reproduction-and-adaptation/>.

<sup>63</sup> Bombay High Court, Ram Sampath v. Rajesh Roshan, 2008 SCC Online Bom 370.

<sup>64</sup> Bombay High Court, Ram Sampath v. Rajesh Roshan, 2008 SCC Online Bom 370.

<sup>65</sup> Sneha Jain and Akshat Agrawal. *Indian Copyright Law and Generative AI: Part 3 – The Output Stage: Analyzing Reproduction and Adaptation*. <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-3-the-output-stage-analyzing-reproduction-and-adaptation/>.

results in an output that is verbatim or substantially similar to the protected expressive aspects of the work, and this output substitutes the original expressive aspects of the market, such use would constitute copyright infringement.

### 3.2.2 The Adaptation Right

The Adaptation Right requires particularly careful analysis in the AI context. While it might initially seem to extend copyright protection to any AI-generated work based on training data, Indian courts have provided crucial limitations and clarifications. The right encompasses uses that alter or rearrange original works while retaining core expressions, but its application to AI transformations demands a specialised understanding.

Furthermore, Indian copyright law does not explicitly grant a statutory right for creating derivative works based on copyrighted material. Indian courts have interpreted the adaptation right under Section 14 to exclude transformative derivative works. This interpretation aligns with the Delhi High Court's approach in *University of Cambridge v. BD Bhandari*, where works with "transformative purposes" or of a "transformative character" were protected.

- **Analysis of Alterations**

The Division Bench of the Calcutta High Court in *Barbara Taylor Bradford v. Sahara Media Entertainment* provides a definitive framework for understanding "alteration" under Section 2(a). The court explicitly rejected an expansive interpretation that would encompass major transformations, stating unequivocally that "*a totally changed thing can never be termed a copy of the original thing.*" The judgment offers specific guidance on interpreting Section 2(a)(v), which defines "adaptation" to include rearrangement or alteration.<sup>66</sup>

The court carefully examined the meaning of "alter," acknowledging its potential to encompass both minor and significant changes. However, it

consciously restricted the term to minor alterations that do not transform the core purpose, character, meaning, or message of the overall work. The court emphasised that attributing a "large change" meaning to Section 2(a)(v) would lead to an absurd interpretation. Instead, it confined the term to "minor change, slight change, not making the original something beyond recognisable possibilities, changes in some of the details."<sup>67</sup>

- **Framework for Evaluating AI Outputs**

For GenAI systems, these principles establish a comprehensive evaluation framework. Merely basing outputs on training datasets arguably does not automatically constitute adaptation right infringement. The test requires examining whether the output is essentially the same work in a different format or merely includes trivial variations. The process of tokenising copyright works for AI training potentially represents a significant transformation that is non-expressive and functional, making it unlikely to substitute the original work's market. Outputs must be assessed for:

1. Degree of transformation from original works
2. Purpose and character of the changes
3. Market substitution potential
4. Nature of similarities (incidental versus substantial)
5. Evidence of mere regurgitation versus genuine transformation

### 3.2.3 Practical Application to AI Systems

These legal principles suggest that GenAI outputs, which substantially transform their inputs rather than making minor alterations, may avoid infringing adaptation rights. The framework established by Indian courts allows for the innovative use of existing works while safeguarding copyright holders' legitimate interests in their original expressions. This interpretation is particularly relevant for AI systems that process and transform training data into new, creative outputs.

<sup>66</sup> Government of India. (1957). *The Copyright Act, 1957 Section 2(a)(v)*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>67</sup> Calcutta High Court. *Barbara Taylor Bradford and Anr. v. Sahara Media Entertainment Ltd. And Ors.*, 2003 SCC Online Cal 323. Also at <https://indiankanoon.org/doc/757852/>

The comprehensive framework emerging from Indian jurisprudence suggests that GenAI systems can operate within copyright law's boundaries when they:

1. Genuinely transform works rather than merely reproducing them
2. Create outputs that serve different purposes from original works
3. Generate content that demonstrates creativity beyond simple arrangement
4. Process training data in ways that produce functional rather than expressive transformations
5. Avoid direct regurgitation or market substitution

Thus, GenAI systems must be designed with specific safeguards, serving as guardrails:

1. Mechanisms to prevent direct regurgitation of training data
2. Processes ensuring substantial transformation of inputs
3. Functions that generate outputs with distinct purposes from original works
4. Systems demonstrating genuine creativity beyond mere arrangement
5. Controls preventing market substitution effects

### 3.3 LIABILITY FRAMEWORKS

#### 3.3.1 Developer Risk Mitigation Framework

A crucial defence against permanent injunctions for GenAI model developers stems from comprehensive risk mitigation mechanisms.<sup>68</sup> Developers who implement all *technologically possible, practically feasible, and objectively plausible steps* to prevent regurgitation of substantially similar or copyright-infringing output may establish an arguably strong defence against liability.<sup>69</sup> This protection particularly applies in scenarios where the display, communication, or publication of potentially infringing content lacks "volitional" conduct or where no "reasonable ground of belief" exists regarding its infringing nature. The implementation of output or copyright filters becomes essential to ensure generated content does not constitute an identical or substantially similar copy of copyright-protected training materials.

#### • Direct and Indirect Liability Analysis

Section 51(a)(i) of the Act, while establishing strict liability, crucially requires an element of "causation" or "volitional conduct," demanding authorisation or control. Developers may avoid direct liability by demonstrating they neither authorised nor permitted regurgitation of substantially similar expressions. For indirect infringement under Section 51(a)(ii), the defence hinges on proving a lack of awareness or reasonable belief of infringement.<sup>70</sup> The Delhi High Court's interpretation in *My Space v. Super Cassettes* defines "reasonable ground for believing" as requiring actual consciousness or awareness, not merely the possibility or suspicion of infringement. This aligns with other legal contexts where the mere possibility of harm, without factual knowledge, is insufficient for establishing liability. However, this interpretation, primarily made in the intermediary context and under Section 79 of the Information Technology Act's safe harbor provisions, remains untested in cases where GenAI models may act as publishers.<sup>71</sup>

#### • User Interaction and Liability

Given prompt engineering's development into a possibly copyright-protected creative activity, the user's contribution becomes crucial in liability assessment. Specific instructions and user interactions help to properly guide model outputs, hence turning simple potential for infringement into real content. However, assigning liability solely to end users raises important considerations. Most users lack the skills to evaluate intellectual property threats or grasp how artificial intelligence models run behind the scenes; they are neither legal nor technical specialists. Holding all users equally responsible could put an excessive burden on individuals who are only using a service that is advertised as secure, even though some users may purposefully manipulate AI tools to create unlawful content—for example, by using prompt injections intended to get over system safeguards.

A more balanced approach requires distinguishing between intentional misuse and responsible use. Clear thresholds should be established to

<sup>68</sup> Katherine Lee, James Grimmelman, A. Feder Cooper, "Talkin' Bout AI Generation: Copyright and the Generative AI Supply Chain", Journal of the Copyright Society of the United States (forthcoming 2024) <https://arxiv.org/pdf/2309.08133>

<sup>69</sup> Sneha Jain and Akshat Agrawal, *Indian Copyright Law and Generative AI: Part 4 – Who is liable for infringing outputs?* <https://www.saikrishnaassociates.com/indian-copyright-law-and-generative-ai-part-4-who-is-liable-for-infringing-outputs/>.

<sup>70</sup> Government of India. (1957). *The Copyright Act, 1957 Section 51*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

<sup>71</sup> Government of India. *Information Technology Act, 2000 Section 79*. [https://www.indiacode.nic.in/bitstream/123456789/13116/1/it\\_act\\_2000\\_updated.pdf](https://www.indiacode.nic.in/bitstream/123456789/13116/1/it_act_2000_updated.pdf).



determine liability, considering factors such as the level of transparency provided to users, the extent of their manipulation of the tool, and their reasonable expectations regarding its legal use. In cases where users intentionally violate terms and conditions to produce prohibited content, they should be held accountable. However, where lack of transparency or system vulnerabilities contribute to misuse, liability may be more appropriately attributed to the developers or organisations responsible for overseeing AI deployment.

- **Comprehensive Protection Measures**

Developers can fortify their position by establishing clear user terms and conditions that explicitly prohibit prompt injections or any manipulation of the model intended to generate substantially similar outputs. These actions should be classified as illegal and outside the scope of permissible use, with liability resting solely on the users. Documenting these comprehensive prevention efforts will showcase the developers' commitment to preventing copyright infringement. When models produce regurgitations even with

### 3.4.2 Strategic Attribution Implementation

A well-designed attribution strategy, particularly for model outputs, significantly strengthens defences against claims of reckless disregard for truth. This strategy involves ensuring that generated content is traceable to credible, well-cited sources and that unreliable or uncredited materials are systematically filtered out during training. Such measures mitigate legal concerns, especially in cases where false information or misattribution could lead to defamation or moral rights claims. Additionally, it enhances the credibility and reliability of the model.

An efficient attribution technique in a legal research AI tool, for instance, can entail directly connecting case law summaries to official court decisions, making it evident if an interpretation is based on secondary comments or primary legal sources. Similar to footnotes in journalism, attribution in an AI system that generates news could entail giving factual references so that readers can confirm the information's original source. By incorporating these approaches, AI developers can increase the reliability of AI-generated information and strengthen moral and legal accountability.

## 3.4 MORAL RIGHTS FRAMEWORK

### 3.4.1 Attribution Requirements

Copyright owners may assert moral rights to attribution under Section 57 of the Act for works used in model training.<sup>72</sup> However, these risks remain minimal when outputs do not show substantial similarity to the training data. The key argument is that mere use of works for model training, without human exposure to the expressive forms or without identifiable original work in the outputs, may not require attribution.

---

<sup>72</sup> Government of India. (1957). *The Copyright Act, 1957 Section 57*. <https://www.copyright.gov.in/Documents/Copyrightrules1957.pdf>

## 4. CONCLUSION

GenAI technology necessitates a reimagining of copyright law to address its unique challenges and opportunities. Existing legal frameworks, designed for an analogue and early digital age, must evolve to account for the complexities of AI systems that learn from and generate content based on copyrighted works. Striking a balance between protecting the rights and investments of creators and fostering innovation is paramount. This calls for nuanced legal provisions that delineate the permissible scope of data use for AI training while safeguarding against potential misuse of expressive works. By creating a robust yet flexible regulatory environment, policymakers can ensure that copyright law remains relevant in the GenAI era—one that not only preserves the integrity of intellectual property but also supports technological advancements that benefit society at large.



# Authors



## Jameela Sahiba

**Associate Director- AI and Public Affairs, The Dialogue**

Jameela Sahiba manages the Secretariat of the Coalition for the Responsible Evolution of AI (CoRE-AI) and leads The Dialogue's work on Artificial Intelligence and public affairs. She coordinates engagement with Parliamentarians and government stakeholders on technology policy. She holds a law degree from Symbiosis Law School, Pune. Prior to her current role, she served as Chief of Staff to Dr. Amar Patnaik, Member of Parliament (Rajya Sabha), where she worked on legislative matters and stakeholder outreach. She has over five years of experience in public policy, with a focus on emerging technologies and legislative processes.



## Sneha Jain

**Partner, Saikrishna & Associates**

Sneha Jain is a Partner at Saikrishna & Associates and a member of the firm's Litigation practice. She advises on litigation and dispute resolution across all areas of intellectual property law, with experience in cross-border IP disputes, multi-jurisdictional arbitrations, media and entertainment, sports broadcasting, defamation, pharmaceutical patent infringement, trademark, data privacy, intermediary liability, online piracy, and competition law. She regularly appears before the High Courts, the Supreme Court of India, sectoral tribunals including TDSAT, and in institutional and ad-hoc arbitrations administered by SIAC and DIAC.



## Akshat Agrawal

**Senior Associate, Saikrishna & Associates**

Akshat Agrawal is a member of the Dispute Resolution and Advisory Practice at Saikrishna & Associates. His work focuses on copyright law, data privacy, music law, artificial intelligence, entertainment and media law, and digital competition. He has advised clients on these issues and regularly appears before courts and tribunals across India, particularly in New Delhi. He is a recipient of Google's Global Award for the best inclusive thesis on copyright law (2023) and has previously served as a fellow with the Wikimedia Foundation.



## Mishthi Dubey

**Senior Associate, Saikrishna & Associates**

Mishthi Dubey is a Senior Associate at Saikrishna & Associates. She works on both contentious and non-contentious intellectual property matters, including patent and trademark litigation, licensing, and technology advisory. She has advised non-profits, artists, and inventors on IP-related issues. She is a member of the firm's generative AI group, which provides risk analysis and mitigation advice to AI developers and technology companies. Mishthi holds a law degree from Gujarat National Law University and an LL.M. from Columbia University. She is enrolled with the Bar Council of India.



[core-ai.in](https://core-ai.in)



[@Core-Ai](https://twitter.com/Core-Ai)



[@CoalitionforResponsibleEvolutionofAI](https://www.linkedin.com/company/CoalitionforResponsibleEvolutionofAI)



[thedialogue.co](https://thedialogue.co)



[@\\_DialogueIndia](https://twitter.com/_DialogueIndia)



[@TheDialogue\\_Official](https://www.instagram.com/TheDialogue_Official)



[@The-Dialogue-India](https://www.linkedin.com/company/@The-Dialogue-India)



[@TheDialogue](https://www.whatsapp.com/channel/01234567890)