ESTD:1989
S.S. RANA & CO.
INTELLECTUAL PROPERTY & CORPORATE LAWS

FICCI 100
100 YEARS OF BUILDING TOMORROW

COPYRIGHT IN THE AI AGE

Protecting Creativity and the Future of Publishing



Copyright © 2026 Federation of Indian Chambers of Commerce and Industry and S.S. Rana & Co.

All contents in this book are the property of the Federation of Indian Chambers of Commerce & Industry and S.S. Rana & Co. and are protected by Indian and international copyright laws. Reproduction, modification, distribution, transmission, republication, display, or performance of any content in this book, in whole or in part, is strictly prohibited without prior written permission from either or both parties.

Copyright in the AI Age: Protecting Creativity and the Future of Publishing

Published in August 2026

Researched and authored by:

Vikrant Rana
Managing Partner, S.S.Rana & Co.

Lucy Rana
Managing Partner, S.S.Rana & Co.

Sumeet Gupta
Deputy Secretary General, FICCI

Gaurav Gaur
Director, FICCI

Aastha Gupta
Senior Assistant Director, FICCI

Assisted in research by:

Pranit Biswas
Managing Associate Advocate, S.S.Rana & Co.

Nikita Sinha
Managing Associate Advocate, S.S.Rana & Co.

Shantam Sharma
Senior Associate Advocate, S.S.Rana & Co.

Huda Jafri
Associate Advocate, S.S.Rana & Co.



Disclaimer

This book examines the legal aspects of copyright, artificial intelligence, and research integrity in the publishing sector, with particular attention to the evolving challenges posed by generative AI and the policy responses needed to address them. The copyright holders have made every effort to secure permissions for the facts, figures, quotations, case references, and other materials reproduced in this book.

This publication is intended solely for informational and educational purposes. It is not intended for any commercial purpose and should not be construed as legal advice, legal opinion, or a substitute for obtaining professional legal counsel.

While utmost care has been taken to ensure the information presented is accurate and current, FICCI and S.S. Rana & Co. make no representation as to its completeness or correctness, particularly regarding ongoing litigation, evolving regulatory frameworks, and case law that may change after publication. FICCI and S.S. Rana & Co. accept no liability for any direct or consequential loss arising from the use of this book or its contents. References to case law, statutes, and regulatory developments are included for informational and educational purposes only, and do not constitute legal advice. Logos, images, and graphics appearing in this book are used solely for illustrative and educational purposes; the parties claim no ownership of, or association with, these logos or images.

Table of Contents

Chapter	Topic	Page
1.	**Introduction: Copyright, AI, and the Publishing Sector** • Background: Convergence of AI, Copyright, and Publishing • Scope, Objectives, and Methodology • Copyright Law Fundamentals • The Generative AI Landscape Relevant to Publishing	**8-12**
2.	**AI Training Data and Copyright Infringement** • Text and Data Mining (TDM): Legal Status Across Jurisdictions • Web Scraping of Copyrighted Works and the "Input" Problem • Fair Use/Fair Dealing Defences: Comparative Analysis (US, EU, UK, and India) • Landmark Litigation and Ongoing Disputes • Licensing Solutions: Collective Licensing, Marketplaces, and Opt-Out Mechanisms • The EU AI Act's TDM Transparency and Opt-Out Regime	**13-18**
3.	**AI-Generated Output: Copyrightability, Authorship, and Ownership** • AI-Generated Copyright: Comparative Positions • Disclosure and Attribution Requirements • Implications for Publishers' Acquisition and Licensing Agreements • A Practical Record-Keeping Framework for Evidencing Human Authorship	**19-26**
4.	**Research Integrity in the Age of AI** • Defining Research Integrity and Research Misconduct • Traditional Threats: Plagiarism, Data Fabrication, and Falsification • Plagiarism in the AI Era • New AI-Enabled Threats	**27-36**
5.	**Regulatory Frameworks and Sector Impact in India** • Indian Legal and Policy Framework	**37-43**
6.	**Contracts, Liability, and Dispute Resolution** • Drafting AI Clauses in Publishing and Author Agreements • Representations, Warranties, and Indemnities Relating to AI Use • Licensing AI Training Rights: Model Contractual Clauses • Royalty and Revenue-Sharing Models for AI-Licensed Content • Determining Liability: Developers, Deployers, and Publishers • Role of Collecting Societies (e.g., IPRS and CRRC) in the AI Era	**44-58**
7.	**Policy Recommendations and the Road Ahead**	**59-63**

Foreword



The publishing industry has always been at the forefront of preserving, disseminating, and advancing knowledge. Today, as Artificial Intelligence rapidly transforms the way content is created, discovered, and consumed, publishers find themselves at the centre of one of the most significant technological shifts of our time. While AI presents immense opportunities to enhance productivity, accessibility, multilingual publishing, and innovation, it also raises fundamental questions around copyright, authorship, licensing, transparency, and research integrity.

For the publishing ecosystem, these are not merely legal or technological issues—they go to the heart of sustaining creativity, rewarding investment in quality content, and maintaining trust in the knowledge economy. As AI continues to evolve, it is essential that innovation and intellectual property protection progress together through balanced, practical, and future-oriented policy frameworks.

Recognizing the importance of these developments, FICCI, in collaboration with S.S. Rana & Co., has undertaken this report to examine the evolving intersection of AI, copyright, and publishing. Drawing upon comparative legal developments, emerging policy frameworks, landmark judicial decisions, and industry practices, the report provides a comprehensive analysis of the opportunities and challenges that AI presents for publishers, authors, researchers, technology developers, and policymakers.

I hope this report serves as a valuable resource for policymakers, publishers, authors, legal professionals, researchers, and technology stakeholders and contributes to informed discussions as well as collaborative action towards building a resilient, responsible, and future-ready publishing ecosystem.

I am confident that this joint initiative will support meaningful dialogue and informed policymaking as the publishing industry navigates the opportunities and challenges of the AI era.

Monica Malhotra Kandhari

Chair, FICCI Publishing Committee

Foreword



In the last decade, the world of creativity and knowledge production has witnessed a transformation more profound than perhaps any in living memory. Generative artificial intelligence has changed how we write, research, and publish, and in doing so, has begun to question the very foundations of copyright law: originality, authorship, and human creative expression.

I still remember the first time I asked an AI tool to summarize a dense legal judgment. Within seconds, it produced a crisp summary that would ordinarily take a junior associate an afternoon to prepare. It was a striking demonstration of capability, and also a moment that raised important questions. Where had this summary come from? Whose words had been learned, absorbed, and repackaged to produce it? And if published as one's own work, whose creativity would it truly represent? These questions sit at the heart of this book.

As a firm that has spent decades helping authors, publishers, and creators protect their intellectual labor, we have watched courts, regulators, and industry bodies grapple with questions no copyright statute anticipated. Can a machine be an author? Can a work made without meaningful human creative control be protected at all? What happens when training data is drawn, without consent or compensation, from the works of those who have spent lifetimes honing their craft?

These are not academic questions. They carry real consequences for authors, publishers, and researchers everywhere, navigating a landscape where old certainties are tested daily by new realities, even as fabricated citations, undisclosed AI-generated content, and paper mills threaten the trust underpinning scholarly work.

This book offers a balanced analysis of these questions, and a ready reference for publishers, authors, researchers, and policymakers alike.

Vikrant Rana

Managing Partner

S.S. Rana & Co.

Chapter I:

Introduction: Copyright, AI, and the Publishing Sector



1.1 Background: Convergence of AI, Copyright, and Publishing

The growth and increasing sophistication of artificial intelligence is reshaping the publishing sector, from content development to distribution, marketing, and reader engagement.

The editorial function benefits from AI-assisted editing and writing-style analysis that improves consistency, while the marketing function benefits from improved book discoverability through personalised recommendations and audience segmentation.

The publishing sector occupies a central position in the debate surrounding generative AI because large language models are trained substantially on text, which constitutes the publishing industry's core intellectual asset. Publishers, news organisations, and authors' associations have consequently faced an unusual dual challenge since the public release of widely adopted generative AI systems in late 2022.

On one hand, their published works are treated as indispensable sources of training data for AI models. On the other hand, these same organisations are increasingly using the same technologies to support editorial assistance, translation, and reader analytics, improving their own productivity. This book refers to this dual position as "double exposure".

This duality raises significant legal, ethical, and commercial questions. Substantial investment goes into the creativity, quality assessment, and editing that underlie published works, and the use of that content without explicit consent or remuneration raises concerns about copyright infringement, the adequacy of existing licensing practices, and the fair valuation of that investment. In response, a number of publishing houses are seeking stronger legal protection, collective licensing arrangements, and greater transparency regarding the use of their content for AI training.

Artificial intelligence offers considerable opportunities for innovation across the publishing sector. At the same time, overreliance on AI without adequate human oversight may compromise accuracy and editorial integrity.

Neither side of this dual exposure is yet settled under existing legal frameworks. It is this unsettled position that this book seeks to examine, considering the various dimensions of the issue and the policy responses it now requires.

Commercial responses have already begun to emerge alongside the litigation discussed in Chapter 2: direct licensing agreements between AI developers and major news and publishing organizations have been reported since 2023–2024, suggesting that the market is moving towards negotiated licensing even as the underlying legal questions remain contested in court. This book treats both tracks, litigation and licensing, as complementary lenses on the same underlying problem

Viewed from an industry and policy perspective, this convergence presents the publishing ecosystem with opportunities as well as challenges. Generative AI can expand access to knowledge, support new editorial and translation workflows, and open licensing revenue that did not previously exist, while also raising legitimate concerns about fair compensation, consent, and the long-term sustainability of original content creation. Government, industry bodies, publishers, technology developers, academia, and creators each have a stake in how this balance is struck, and the discussion in the chapters that follow is intended to inform that shared conversation rather than to advocate for any single outcome.

1.2 Scope, Objectives, and Methodology

This book's scope is deliberately bounded to text-based copyright works of direct relevance to publishing: trade books, academic and STM journals, news content, and reference works. Image- and music-generation disputes (e.g., visual-art AI litigation) are referenced only comparatively, where a doctrinal point translates cleanly across media.

Three objectives help in structuring the analysis: *first*, to map the legal exposure created by AI training and AI-generated output for publishers operating across multiple jurisdictions; *second*, to diagnose the research-integrity risks that generative AI introduces into academic and scholarly publishing, a distinct but related problem from copyright infringement; and *third*, to translate both analyses into concrete contractual clauses and policy recommendations that a publisher, in-house counsel, or external IP advisor can use directly.

Methodologically, the book proceeds by comparing statutory analysis across the United States, United Kingdom, European Union, and India; by tracking landmark and ongoing litigation as a live record of how courts are testing existing doctrine against AI facts; and by practitioner-oriented drafting guidance drawn from the authors' transactional experience.

1.3 Copyright Law Fundamentals

1.3.1 Originality, Authorship, and Ownership

The foundation of copyright across jurisdictions is based on the two interrelated factors: whether the work is original enough to be eligible for copyright protection and who can be recognized as its author. These two factors are equally relevant in this age of generative artificial intelligence also where the main challenge is to ascertain distinction between human and machine generated work.

Even though ‘originality’ is the essential pre-requisite for copyright protection, versatility in its interpretation has been observed across jurisdictions due to the legal standards in its assessment. The jurisdictions compared in this book answer the originality question differently in emphasis, though the results often converge.

In the United States of America, “modicum of creativity” governs the originality of a work. This standard was established by the Supreme Court in *Feist Publications, Inc. v. Rural Telephone Service Co.*[1]. The Court rejected the ‘sweat of the brow’ doctrine which traditionally allowed the copyright primarily based on ‘effort, labor and judgement’ invested in creation of the work i.e., the factual compilation in this case.

Instead, the Court held that the independently created work of an author, comprising at least a minimal degree of creative expression, is eligible for copyright protection. This case has a direct influence on the present scenario being discussed in this book as AI systems are majorly trained on a vast collection of factual as well as creative information. The concept of authorship is based on the presumption that the human author has exercised intellectual judgement and creative control over the work. As the generative AI is continuously restructuring the production of creative works, its ownership, protection and commercial exploitation are the factors on which the legal debates are revolving around, concerning the principles of originality and authorship.

[1] *Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991).*

The United Kingdom and India, by contrast, historically applied the different standard of 'skill, labour and capital' articulated through the judgement in *University of London Press v. University Tutorial Press*[2]. This approach can be said to be a lower threshold than the US standard as it emphasizes primarily on the process of creation and independent effort of author rather than the degree of creativity. As a result, the directories and compilations were found to be eligible for copyright protection.

India witnessed a significant development through the Supreme Court judgement of *Eastern Book Company v. D.B. Modak*[3] where the Court adopted a higher threshold requiring the application of skill and judgment involving a minimal degree of creativity in ascertaining the originality of the work i.e., the editorial version of the court judgements.

With time, the position in India and UK has been influenced and evolved by the approach of the European Union Court which focused on the concept of author's 'own intellectual creation' which shifted the focus to creative choices of the author rather than mere labor and investment. The AI generated content is tested on this very ground as to whether the expression of work reflects human's creative choices, rather than a machine's statistical output.

These differing formulations converge on one point critical to this book: every regime surveyed ties originality, at least implicitly, to a human creative act. That convergence is precisely the fault line AI strains, and it recurs throughout Chapters 2 and 3.

1.3.2 Economic and Moral Rights

The economic rights most directly affected by the generative AI is a combination of rights of reproduction, adaptation, distribution and communication to the public. The content generated by AI is primarily based on the reproduction of already available content from books, journals, websites and converting them into machine readable format. This may also amount to unauthorized access to the original work undermining commercial interests of copyright holders.

Additionally moral rights represent another significant aspect of copyright law, principally the rights of attribution and integrity, recognized in Article 6bis of the Berne Convention. These rights receive stronger legal protection in India as well as United Kingdom and European Union.

Since the AI based content uses, modifies and sometimes distorts the author's work that too without providing renumeration, the question regarding the integrity, lack of attribution directly implicates the moral rights of author.

The Courts and legislatures at present face challenges in addressing protection of both economic and moral rights along with balancing innovation through AI generated content.

1.3.3 Fair Dealing/Fair Use Exceptions

In the US, the fair use doctrine codified at 17 U.S.C. § 107[4] is based on the following four factors:

(i) the purpose and character of the use (including whether it is transformative),
(ii) the nature of copyrighted work,

[2] *University of London Press v. University Tutorial Press [1916] 2 Ch 601.*
[3] *Eastern Book Company v. D.B. Modak (2008) AIR 2008 SC 809.*
[4] https://www.copyright.gov/title17/92chap1.html#107

(iii) the amount and substantiality of the portion used in relation to the copyrighted work
(iv) the effect of the use on the potential market for the original work

These standards primarily discussed for training AI models are covered in Chapter 2.3 of this book.

Out of the four factors stated above, the concept of transformative use is widely discussed from the aspect of AI related contents as the developers contest in defense that content created is result of training large language models and analysis of concepts and thus is a transformative work.

On the other hand, copyright owners argue that the use of protected work for commercial AI development crosses the border of legitimate scope of fair use.

The UK and India again follow a narrow approach in determining its scope and instead allows the use of copyrightable work for specific purposes laid down by the legislations including research, review, education, judicial proceedings.

As the concept of fair use was adopted far before the emergence of AI, the AI developers face a large uncertainty in scope of its coverage. Considering that AI continues to transform the creation of work, the practical implementation of fair use and fair dealing is likely to be a point of legal debate across jurisdictions.

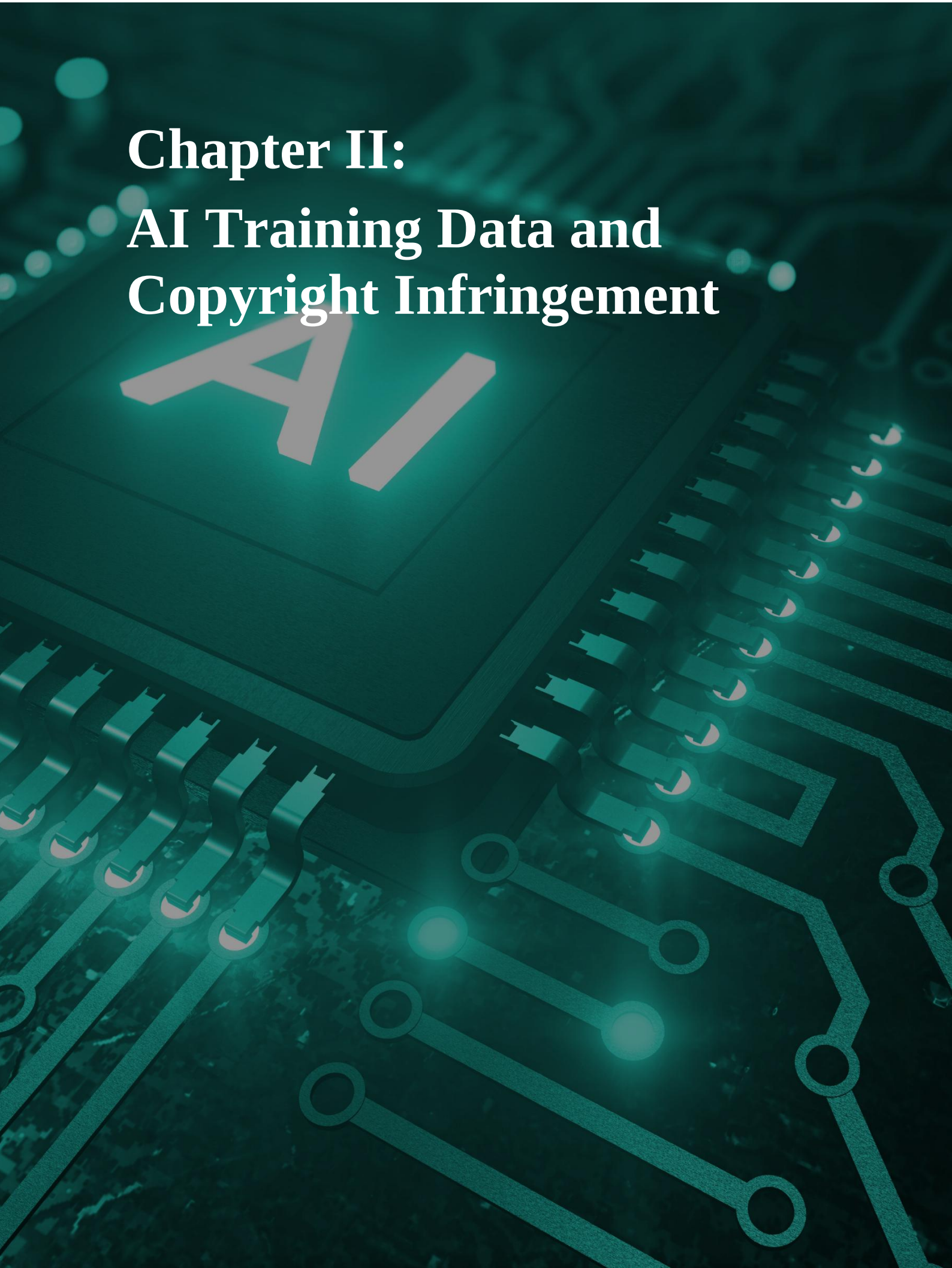
1.4 The Generative AI Landscape Relevant to Publishing

It is imperative for the readers to understand the basics of how generative AI functions, either from publishing or legal point of view. LLM (large language models) are designed to understand and create human language by analyzing huge collections of text and datasets enabling it to predict the probable sequence of text depending on its training process. The training involves multiple inter-related stages including data collection and curation, tokenization (where texts are converted to small units/ tokens). Although this process appears to be highly technical but as it involves reproduction of copyrighted work in digital medium, it raises concerns regarding the owner's reproduction rights.

Two technical characteristics of LLM have legal relevance: (i) memorization and (ii) output similarity discussed further in Chapter 2.4.

Memorization generally occurs in case the training data contains distinctive and repeated texts. The right holders in contentious matters relied on such instances as proof that the copied content is retained within the AI systems during the training process.

The characteristic of output similarity deals with the fact that even without the verbatim memorization, the output can still have a close resemblance in pattern, style and narration of the training works due to statical pattern from the copyrighted content which is more complex from the infringement perspective.



Chapter II:
AI Training Data and Copyright Infringement

Generative AI systems learn to write, draw, or code by processing huge amounts of existing material, most of it protected by copyright. This chapter looks at how the law in India and the United States, the European Union, the United Kingdom, treats this process, known as text and data mining (TDM), before turning to the court decisions that have begun to grapple with the central question: does training a model on copyrighted works without a licence amount to infringement?

2.1 Text and Data Mining

Text and data mining refers to the automated analysis of large volumes of digital material to extract patterns or other information. The European Union was the first major jurisdiction to legislate on this directly, through the Directive on Copyright in the Digital Single Market. Article 3 allows TDM for scientific research by research organisations and cultural heritage institutions, with no opt-out available to rightsholders. Article 4 allows TDM for any purpose, including commercial AI development, but rightsholders can reserve their rights, and once they do, the exception stops applying to that work. For content made publicly available online, that reservation must be expressed through machine-readable means; for other content, rightsholders may reserve their rights through other appropriate means, such as contractual agreements or a unilateral declaration.[5]

The United Kingdom took a narrower approach. Section 29A of the Copyright, Designs and Patents Act 1988, added in 2014, allows TDM only for non-commercial research, and only where the person doing the mining already has lawful access to the material. There is no UK exception for commercial TDM, so training a commercial large language model on scraped text falls outside it.[6]

The UK Intellectual Property Office consulted on an EU-style opt-out exception between December 2024 and February 2025, but the Government rejected that approach in its March 2026 report and left the existing copyright framework unchanged. Instead, the Data (Use and Access) Act 2025 required the Government to publish reports on issues such as transparency, licensing, and technical standards for AI, while deferring any substantive legislative reform.[7]

The United States has no dedicated TDM exception. Whether the copying involved in building a training corpus, and in the training process itself, is lawful depends on the four-factor fair use test in Section 107 of the Copyright Act, applied case by case by federal courts.

India has no TDM provision either. Section 52 of the Copyright Act, 1957[8] sets out a closed list of fair dealing exceptions, including one for private or personal use, including research. Whether AI training fits within that category was untested until 2026, when the Delhi High Court addressed it for the first time.[9]

The absence of an opt-out in the US, UK, and Indian columns does not mean rightsholders have no protection there. It means those systems do not use the EU's specific mechanism of a default permission that can be individually withdrawn. In the US, fair use is open-ended and case-specific. In India, Section 52 contains enumerated fair dealing exceptions, applied by

[5] *Directive (EU) 2019/790, Articles 3 and 4.*
[6] *Copyright, Designs and Patents Act 1988, s. 29A.*
[7] *https://www.gov.uk/government/publications/report-and-impact-assessment-on-copyright-and-artificial-intelligence/report-on-copyright-and-artificial-intelligence?*
[8] *The Copyright Act, 1957, s. 52(1)(a).*
[9] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024.*

courts on a case-specific basis within the statutory list. In the UK, it happens because the exception is written narrowly enough that most commercial training falls outside it.

2.2 Web Scraping and the "Input" Problem

Before a model can be trained, its developer needs a dataset, which in practice has meant scraping billions of pages of text and millions of images from the open web, sometimes supplemented by digitised books and licensed archives. This raises what is known as the input problem: even if a model's output is transformative, the copying, downloading, and storing that happens before training may itself count as reproduction under copyright law, separate from anything the model later produces.

In *Bartz v Anthropic*[10], the Northern District of California drew a key distinction, not between lawfully acquired and pirated books, but between different uses of the same books. Training Claude's underlying models on copyrighted books, purchased or pirated, was fair use either way. What was not fair use was something separate: Anthropic had downloaded more than seven million books from shadow libraries such as LibGen and the Pirate Library Mirror to build a permanent digital library, and kept them there whether or not they were ever used for training. The court held that this hoarding was not protected by fair use.

Robots.txt files and other machine-readable opt-out signals have become a practical point of dispute, particularly in Europe. In the *LAION dataset case*[11], a German photographer sued a non-profit that had catalogued one of his images within a five-billion-image training dataset. The Hamburg court dismissed the claim on the basis that LAION had only assembled the dataset and had not trained a model on it, so the copying fell within the non-commercial scientific research exception, which carries no opt-out, rather than the commercial TDM exception that does.

The court went on to consider Article 4(3)'s commercial opt-out mechanism. It suggested that the photographer's prohibition on AI use, contained in the website's terms and conditions, could potentially satisfy the requirement for a machine-readable rights reservation. The court reasoned that, as technology develops, AI systems may be capable of automatically recognising and processing rights reservations written in ordinary language, rather than only technical signals such as robots.txt or metadata. The Hanseatic Higher Regional Court dismissed the photographer's appeal on 10 December 2025. The court went further than the lower court, expressly holding that a natural-language opt-out did not satisfy the statutory machine-readability requirement, and that LAION could rely on the text and data mining exception, while permitting a further appeal to Germany's Federal Court of Justice.[12]

2.3 Fair Use and Fair Dealing: US, EU, UK, and India

2.3.1 United States

Fair use in the US turns on four factors: the purpose and character of the use, the nature of the copyrighted work, the amount used relative to the whole, and the effect on the market for the original. The Supreme Court's 2023 decision in *Andy Warhol Foundation v Goldsmith* narrowed [13]

[10] *Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025).*
[11] *Kneschke v. LAION e.V., Landgericht Hamburg, Case No. 310 O 227/23.*
[12] *Kneschke v. LAION e.V., Hanseatisches Oberlandesgericht Hamburg, Case No. 5 U 104/24.*
[13] *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023).*

In *Bartz v Anthropic*, the Northern District of California described training an LLM on the books in Anthropic's library, whatever their source, as "quintessentially transformative," comparing it to a student learning to write by reading widely. The same court reached a similar result days later in *Kadrey v Meta Platforms*, on summary judgment, while noting that plaintiffs who could show genuine market dilution from an AI system might fare differently at trial.[14]

In *Thomson Reuters v Ross Intelligence*, the District of Delaware held that an AI legal research tool's use of Westlaw's copyrighted headnotes to build a competing legal research product was not fair use because it served substantially the same commercial purpose as Westlaw and competed in the same market. The court distinguished generative AI cases, noting that Ross's system was not an LLM. The decision is now on appeal before the Third Circuit.[15]

2.3.2 European Union

The EU's Article 4 TDM exception works as a default licence: commercial AI training is permitted unless a rightsholder validly reserves its rights. This places the burden on rightsholders to police their own opt-outs, and on AI developers to build systems that can detect and honour those reservations.

2.3.3 United Kingdom

Because Section 29A is limited to non-commercial research, large-scale commercial AI training generally cannot rely on it. A developer training a model on scraped copyrighted text in the UK, without a licence, is infringing.

2.3.4 India

The Delhi High Court's decision in ANI Media Pvt. Ltd. v. Open AI OpCo LLC held, at the interim stage, that OpenAI's storage and use of ANI's copyrighted news content for training LLMs underlying ChatGPT fell within the scope of the "private or personal use, including research" exception under Section 52(1)(a) of the Copyright Act, 1957, and therefore did not amount to prima facie infringement. The Court rejected the argument that "research" must be confined to human activity and observed that machine learning undertaken for human benefit could fall within the provision. It also found that ChatGPT outputs, on the material before it, did not substantially reproduce ANI's works, and that ANI had not established memorisation or regurgitation of its content. The Court nevertheless clarified that its observations were limited to the interim application and would not determine the final outcome of the suit.[16]

2.4 Landmark Litigation and Ongoing Disputes

2.4.1 *Getty Images v Stability AI*

Getty Images v Stability AI began when Getty filed proceedings in the High Court of Justice of England and Wales on 16 January 2023, alleging that scraping and using millions of Getty photographs to train the Stable Diffusion model infringed its copyright, trade marks, and database rights. Judgment was handed down by the High Court on 4 November 2025.[17]

[14] *Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025); Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417 (N.D. Cal. June 25, 2025).*
[15] *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB (D. Del.).*
[16] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024.*
[17] *Getty Images (US) Inc & Ors v. Stability AI Limited [2025] EWHC 2863 (Ch).*

Getty abandoned its primary training and development claim during the trial, after acknowledging it could not show the relevant activity took place within the jurisdiction. What remained was a secondary infringement claim: that Stable Diffusion was an "infringing copy" under section 27 of the CDPA that had been imported into the UK. The court rejected that claim, holding that the model's weights do not store reproductions of the training images. Getty secured a narrow win on trade mark infringement over earlier versions of the model that had reproduced the Getty watermark in generated outputs.[18]

2.4.2 *ANI Media v OpenAI*

ANI Media v OpenAI, filed in November 2024, was the first copyright action brought by an Indian media organisation against OpenAI. ANI alleged that ChatGPT was trained on its published reports without permission, and that the chatbot had on occasion generated fabricated stories falsely attributed to ANI. The Digital News Publishers Association later intervened in the proceedings.[19]

On 24 July 2026, in a 135-page order after 32 hearings, the Delhi High Court refused ANI's application for an interim injunction. The court held, on a prima facie basis, that OpenAI's storageof ANI's literary works for training LLMs underlying ChatGPT fell within the fair dealing exception for private use and research under Section 52(1)(a), and rejected ANI's argument that Parliament intended to exclude commercial entities from that exception. The Court held that "research" under Section 52(1)(a)(i) should not be confined to human activity and could extend to machine learning undertaken for human benefit. The court also found that ANI had not shown that ChatGPT memorised or reproduced its work, or that outputs generated by ChatGPT were substantially similar to ANI's original works, and confirmed Indian territorial jurisdiction over OpenAI. The main suit, including ANI's separate claim about fabricated attribution, remains pending.[20]

2.4.3 Other Global Precedents

Thomson Reuters v Ross Intelligence remains the only US decision to reject an AI fair use defence outright, and it is now on appeal before the Third Circuit.[21]

In *Kneschke v LAION e.V.*, the Hamburg Regional Court dismissed the claim based on the non-commercial scientific research exception, and offered an obiter view that a rights reservation expressed in ordinary language on a website could potentially satisfy the machine-readability requirement for the commercial opt-out exception.

The Hanseatic Higher Regional Court went further than the lower court in December 2025, expressly finding that the natural-language opt-out did not satisfy the machine-readability requirement and that LAION could rely on the text and data mining exception, with a further appeal to Germany's Federal Court of Justice still available.[22]

Bartz v Anthropic, though a domestic US dispute, took on wider significance. The Northern District of California's split summary judgment ruling exposed Anthropic to potentially large statutory damages over the pirated portion of its training library, and the parties reached a preliminary settlement in September 2025. The court granted final approval on 20 July 2026,

[18] *Getty Images (US) Inc & Ors v. Stability AI Limited [2025] EWHC 2863 (Ch).*
[19] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024.*
[20] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024.*
[21] *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 25-2153 (3d Cir.) (appeal pending).*
[22] *Kneschke v. LAION e.V., Landgericht Hamburg, Case No. 310 O 227/23.*

for 1.5 billion US dollars, covering more than 482,000 works at roughly 3,000 dollars per book, the largest copyright recovery in US history. Because the matter settled, it sets no binding appellate precedent, though the court's underlying fair use ruling is still being cited by courts and litigants elsewhere.[23]

2.5 Licensing Solutions: Collective Licensing, Marketplaces, and Opt-Out Mechanisms

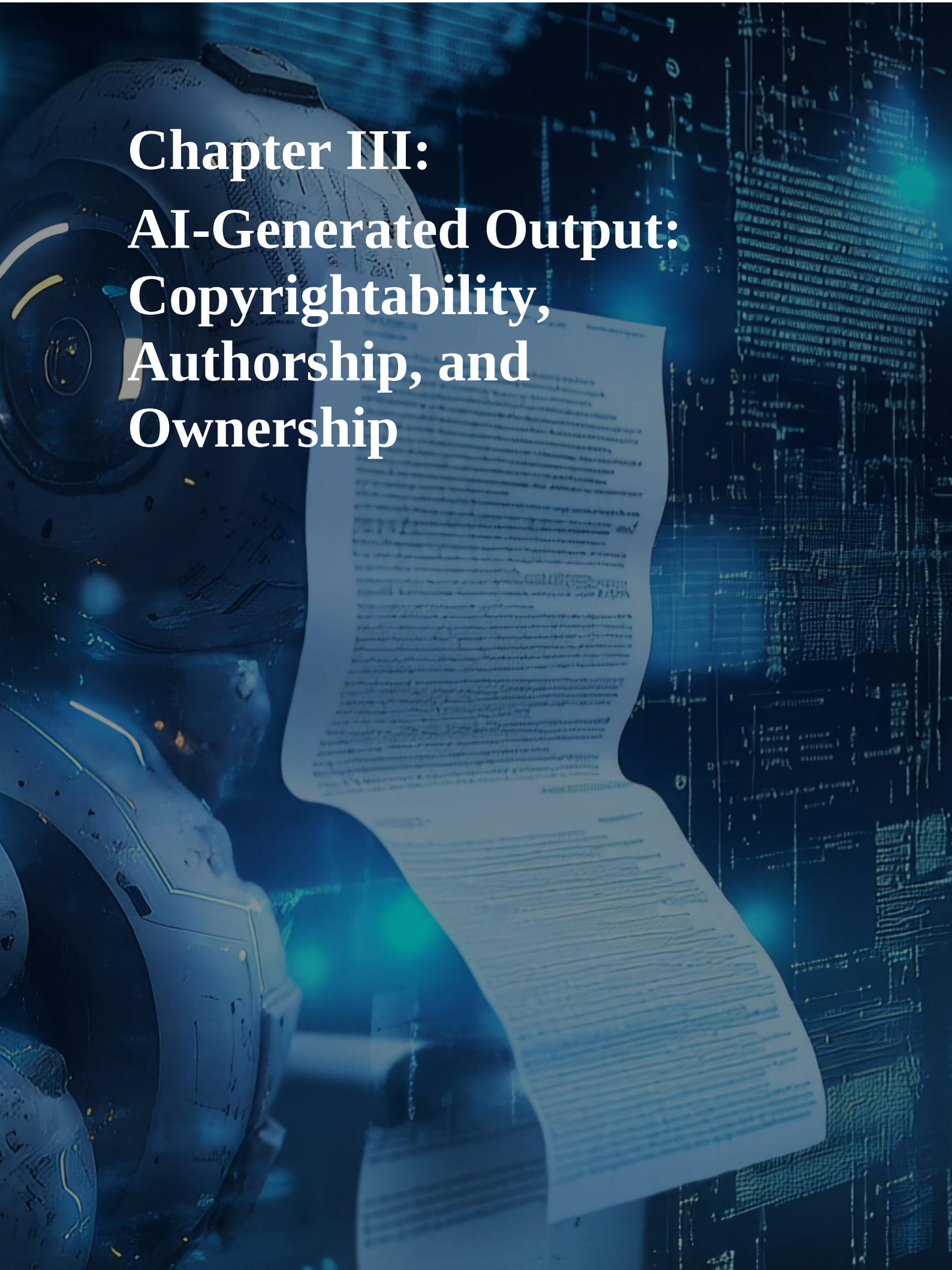
Alongside the litigation, a licensing market for AI training data has emerged. Major publishers, stock image libraries, and news organisations have entered direct licensing agreements with AI developers, trading access to their archives for fees, indemnities, or both.

Collective licensing bodies, long used for music and text reproduction rights, are extending their mandates to cover AI training, offering rightsholders a lower-cost route to compensation than individual negotiation or litigation. A newer category of TDM marketplaces and data licensing platforms has also emerged, intermediating between individual creators and large AI developers.

Opt-out registries and machine-readable reservation protocols, particularly relevant under the EU copyright framework, provide a way for rightsholders to reserve their rights and prevent their works from being used under the commercial TDM exception in Article 4 of the DSM Directive[24]. They operate alongside licensing systems rather than replacing them.[25]

Direct licensing favours large rightsholders with the resources to negotiate, leaving smaller creators dependent on collective bodies that are, in most jurisdictions, still building the infrastructure to represent them. Opt-out systems place the burden of vigilance on rightsholders rather than on those seeking to use their work.

[23] *Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025).*
[24] *https://eur-lex.europa.eu/eli/dir/2019/790/oj*
[25] *https://digital-strategy.ec.europa.eu/en/faqs/stakeholder-consultation-ai-and-copyright-compliance?*



Chapter III:
AI-Generated Output: Copyrightability, Authorship, and Ownership

3.1 AI-Generated Copyright: Comparative Positions

Every jurisdiction surveyed in this chapter asks, at bottom, the same question: does the work owe its existence to a sufficient degree of human creative choice, or did a machine determine its expressive content? Each insists, in one form or another, on human authorship as the touchstone of protection, yet they part company sharply over how that requirement is applied, what evidence is required to demonstrate human contribution, and whether some lesser, residual form of protection should be available for content that no human authored in the conventional sense. The sections that follow set out, in turn, the positions taken in the United States, the United Kingdom, the European Union, and India.

3.1.1 United States

The U.S. Copyright Office set out its governing position in its March 2023 Copyright Registration Guidance on works containing AI-generated material[26], which establishes four principles: human authorship is a prerequisite for registration; copyright does not vest automatically in AI-generated content where the applicant has not exercised creative control; where a work combines human and AI-generated content, protection extends only to the human-authored elements; and applicants must disclose the use of AI-generated content when registering a work that combines the two.

The Office elaborated on that position in Part 2 of its Report on Copyright and Artificial Intelligence, published on 29 January 2025 after considering more than 10,000 public comments[27]. Part 2 reaffirmed that copyright protects only the product of human creativity, declined to recommend either a separate registration track or a sui generis right for AI-generated content, and confirmed that the sufficiency of human authorship would continue to be assessed case by case rather than under a bright-line rule. Where a human author's own copyrightable expression survives in the final work, whether as original text, a selected and arranged sequence of AI outputs, or an image supplied as an input, the Office will register that human contribution notwithstanding the use of AI tools elsewhere in the same work.

The leading authority is *Thaler v. Perlmutter*. Dr. Stephen Thaler sought to register an artwork, 'A Recent Entrance to Paradise', naming his AI system, the 'Creativity Machine', as sole author and describing himself only as the work's owner. The U.S. District Court for the District of Columbia held that a work generated entirely by the system, without any human author, could not be registered on those terms, and the D.C. Circuit affirmed on 18 March 2025[28], holding as a matter of statutory interpretation that the Copyright Act "requires all eligible work to be authored in the first instance by a human being." The Court of Appeals confined its holding to works with no human author whatsoever; it declined to decide what degree of human involvement in an AI-assisted work would suffice for copyrightability, leaving that harder line-drawing exercise to later cases. The Supreme Court denied certiorari on 2 March 2026, letting the D.C. Circuit's ruling stand.[29]

That harder line has since been drawn, at least provisionally, in two administrative decisions concerning AI-assisted, rather than wholly AI-generated, works. In *Zarya of the Dawn*, the Office considered a graphic novel by Kris Kashtanova combining her own text with images

[26] *U.S. Copyright Office, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16,190 (Mar. 16, 2023): https://www.copyright.gov/ai/ai_policy_guidance.pdf.*
[27] *U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability (Jan. 29, 2025): https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf.*
[28] *Thaler v. Perlmutter, No. 23-5233 (D.C. Cir. Mar. 18, 2025), aff'g No. 1:22-cv-01564 (BAH) (D.D.C. Aug. 18, 2023).*
[29] *Thaler v. Perlmutter, No. 25-449 (U.S. cert. denied Mar. 2, 2026).*

generated using Midjourney[30]. It registered Kashtanova's text and her selection, coordination, and arrangement of the AI images alongside that text, but excluded the images themselves. Its reasoning was that a Midjourney user's prompts merely influence rather than dictate the output, since an identical prompt can produce materially different images on each run, so the user does not control the image the way a photographer controls a photograph. The Office's Review Board reached the opposite result on a different point in the *Théâtre D'Opéra Spatial* matter, upholding a refusal to register an award-winning image that Jason Allen had produced through more than 624 iterative Midjourney prompts and further editing in Photoshop and Gigapixel AI, because Allen declined to disclaim the AI-generated portions of his application[31]. Allen has since challenged that refusal before the U.S. District Court for the District of Colorado in Allen v. Perlmutter, No. 1:24-cv-02665 (D. Colo., filed 26 September 2024); the case remained pending at the summary-judgment stage. The dispute illustrates that the line between permissible prompting and impermissible reliance on AI output is still being tested in the courts.

Thaler therefore settles only the easy case: a work with no human author at all cannot be protected. It says nothing about how much human involvement in an AI-assisted work is enough, and that harder question is being worked out, case by case, through registration practice of the kind illustrated by *Zarya of the Dawn* and *Théâtre D'Opéra Spatial*.

3.1.2 United Kingdom

The United Kingdom departs from the United States, the European Union, and most other jurisdictions surveyed here, each of which insists on a human author as a condition of protection. Section 9(3) of the Copyright, Designs and Patents Act 1988 instead deems the person who made the arrangements necessary for the creation of a computer-generated work, one produced in circumstances with no human author, to be its author[32]. The shorter term that attaches to such works, 50 years from creation rather than life of the author plus 70 years, itself signals Parliament's view that computer-generated works merit more limited recognition than works of human authorship.

That position is now under active review. In December 2024, the UK government opened a ten-week consultation on Copyright and Artificial Intelligence, canvassing options for Section 9(3) ranging from retaining it unchanged to narrowing its scope or repealing it outright[33]. Having considered 11,520 responses, 10,110 through Citizen Space and 1,410 by email, the government published its Report on Copyright and Artificial Intelligence on 18 March 2026. The Report proposes removing protection for computer-generated works absent evidence that the provision continues to serve a purpose, while committing to keep its practical impact under review; it stops short, however, of firm proposals on the separate and more contentious question of a commercial text and data mining exception[34]. No legislative timetable has yet been set, and Section 9(3) remains in force for the present, but publishers relying on the 50-year computer-generated-work term for AI-assisted output should treat that basis for protection as provisional rather than settled.

---

[30] *Zarya of the Dawn, Registration No. VAu001480196 (U.S. Copyright Office Cancellation Decision, Feb. 21, 2023).*
[31] *Théâtre D'Opéra Spatial (U.S. Copyright Office Review Board, Sept. 5, 2023).*
[32] *Copyright, Designs and Patents Act 1988, c. 48, § 9(3) (UK): https://www.legislation.gov.uk/ukpga/1988/48/section/9/2003-10-31.*
[33] *Department for Science, Innovation and Technology, Intellectual Property Office & Department for Culture, Media and Sport, Copyright and Artificial Intelligence: Consultation (17 Dec. 2024–25 Feb. 2025): https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence.*
[34] *Department for Science, Innovation and Technology, Report on Copyright and Artificial Intelligence (18 Mar. 2026) (presented pursuant to s.136, Data (Use and Access) Act 2025): https://www.gov.uk/government/publications/report-and-impact-assessment-on-copyright-and-artificial-intelligence/report-on-copyright-and-artificial-intelligence.*

3.1.3 India: Statutory Recognition of Computer-Generated Works

India adopts a broadly similar statutory approach to the United Kingdom's, departing from the position taken in the United States and most other jurisdictions surveyed in this chapter, which generally require an identifiable human author as a precondition of copyright protection. Section 2(d)(vi) of the Copyright Act, 1957 provides that, for a computer-generated literary, dramatic, musical, or artistic work, the author is the person who causes the work to be created.[35] Introduced by the Copyright (Amendment) Act, 1994, this provision mirrors the language of section 9(3) of the UK's Copyright, Designs and Patents Act 1988 discussed in Section 3.1.2 above, and, like its UK counterpart, accommodates works produced without a traditional human author.

3.1.4 European Union

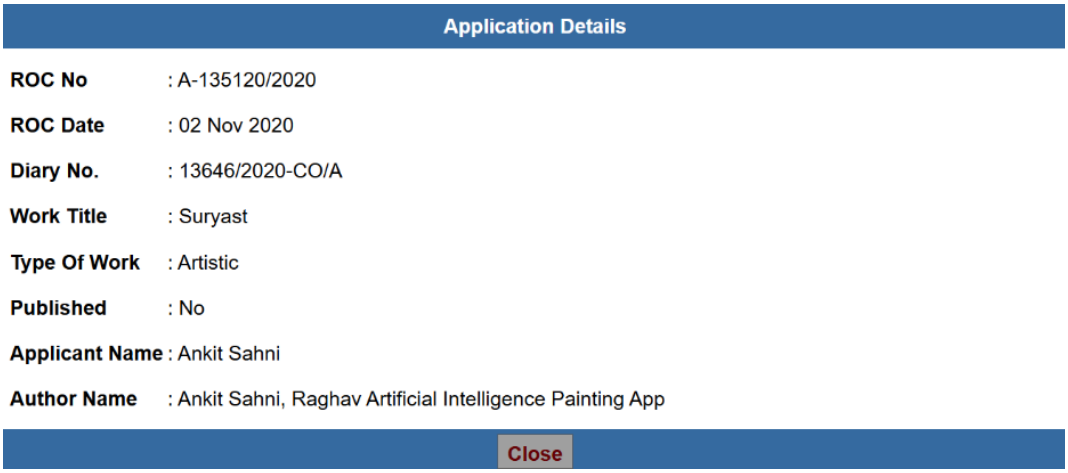
The European Union has no harmonised statutory provision addressing AI-generated or computer-generated works; instead, it assesses originality under the 'author's own intellectual creation' standard, one that becomes difficult to apply whenever the human contribution to the final output falls short of genuine creative control.

That standard traces to the Court of Justice's decision in *Infopaq International A/S v Danske Dagblades Forening*, which held that copyright protects only elements that are 'the expression of the intellectual creation of their author'[36]. The Court elaborated the test in *Painer v Standard VerlagsGmbH*, explaining that intellectual creation requires free and creative choices that stamp the work with the author's personal touch[37]. It has not yet been asked to apply either ruling directly to an AI-generated or AI-assisted output, leaving the question to national courts and copyright offices whose practice remains uneven. A December 2024 policy questionnaire circulated among Member States found a broad, if imprecise, consensus that AI-generated content can attract protection only where the human input into the creative process is significant, a formulation that echoes, without resolving, the same line-drawing difficulty facing the US and UK regimes. A small number of jurisdictions in the wider European area, Ireland among them, retain a UK-style provision for computer-generated works, a reminder that the EU's harmonised originality standard has not produced a common approach to works lacking any human author

3.1.5 The Raghav Application

India's Copyright Office made international headlines in November 2020 when it registered an artistic work titled 'Suryast' (ROC No. A-135120/2020), listing both Ankit Sahni and an AI tool, the Raghav Artificial Intelligence Painting App, known in public reporting simply as 'Raghav', as co-authors[38]. Suryast had been generated by feeding Raghav two inputs, a photograph of a sunset taken by Sahni and Vincent van Gogh's *The Starry Night* as a style reference, together with a variable controlling the intensity of the style transfer.

[35] *Copyright Act, 1957, s. 2(d)(vi), as inserted by the Copyright (Amendment) Act, 1994.*
[36] *Infopaq International A/S v Danske Dagblades Forening, Case C-5/08, EU:C:2009:465 (CJEU, 16 July 2009).*
[37] *Painer v Standard VerlagsGmbH, Case C-145/10, EU:C:2011:798 (CJEU, 1 Dec. 2011).*
[38] *Copyright Office (India), Register of Copyrights entry for 'Suryast', ROC No. A-135120/2020 (Nov. 2020): https://copyright.gov.in/SearchRoc.aspx.*



Application Details

ROC No	: A-135120/2020
ROC Date	: 02 Nov 2020
Diary No.	: 13646/2020-CO/A
Work Title	: Suryast
Type Of Work	: Artistic
Published	: No
Applicant Name	: Ankit Sahni
Author Name	: Ankit Sahni, Raghav Artificial Intelligence Painting App

Close

The registration proved short-lived in its original form, though the underlying question was never actually settled. On 25 November 2021, roughly a year after registration, the Copyright Office issued Sahni a notice proposing to withdraw it, pointing to Sections 2(d)(iii) and 2(d)(vi) of the Copyright Act, 1957, which define the 'author' of an artistic work as the artist or, for a computer-generated work, the person who causes it to be created[39]. Sahni's response was that the Act contains no express mechanism allowing the Registrar to review the Registrar's own prior decision, unlike the equivalent review provisions under the Trade Marks Act and the Patents Act, and no final, reasoned order on the withdrawal appears to have issued since. As of the most recently available Copyright Office records, the Suryast entry, also recorded under Diary No. 13646/2020-CO/A, still shows a status of 'registered', notwithstanding the outstanding withdrawal notice.

Sahni pursued parallel applications for Suryast elsewhere, and the divergent results show how differently comparable facts can fare across systems that each nominally require human authorship. Canada's Intellectual Property Office granted a registration listing Sahni and Raghav as co-authors, but only because its online filing system does not verify authorship particulars, as Canadian authorities have themselves acknowledged; that registration has since been challenged by the Canadian Internet Policy and Public Interest Clinic[40]. The United States went the other way: the Copyright Office refused registration outright, and its Review Board affirmed the refusal on 11 December 2023, holding that the expressive elements of Suryast, the placement, arrangement, and colouring of the pictorial elements, were determined by Raghav rather than Sahni, whose own contribution went no further than selecting the unprotectable idea of applying a chosen style to a chosen photograph[41].

That refusal did not end the matter. In May 2026, Sahni filed a complaint in the United States District Court for the Central District of California, styled Suryast U.S. Enterprises, LLC v. Perlmutter, No. 2:26-cv-04999 (C.D. Cal.), seeking judicial review of the Review Board's decision, arguing that his choice of style, source image, and transfer intensity amounted to more than the Office had credited him with. Like Jason Allen's pending challenge to the *Théâtre D'Opéra Spatial* refusal discussed in Section 3.1.1, this litigation may give a US court its first real opportunity to decide how much curatorial and directional control over a style-transfer or diffusion tool is enough for human authorship.

---

[39] *Copyright Office (India), Notice of Withdrawal re: 'Suryast' (ROC No. A-135120/2020) (25 Nov. 2021), as reported in: https://www.managingip.com/article/2a5bqtj8ume32iwlaoy5y/exclusive-indian-copyright-office-issues-withdrawal-notice-to-ai-co-author.*
[40] *CIPPIC v. Sahni, Federal Court of Canada, Court File No. T-1717-24.*
[41] *SURYAST, Registration Refusal (U.S. Copyright Office Review Board, Dec. 11, 2023).*

A small number of jurisdictions have gone further in recognising AI-assisted authorship. In November 2023, the Beijing Internet Court held in *Li v. Liu* that an image generated using Stable Diffusion was copyrightable, reasoning that the claimant's choice and sequencing of prompts, and his selection among the resulting outputs, reflected enough personalised intellectual investment to satisfy Chinese law's originality requirement[42]. Decided by a court of first instance and carrying no binding precedential weight under Chinese law, the ruling nonetheless sits at the permissive end of a spectrum that otherwise clusters around the stricter positions taken in the United States and India.

The Suryast registration illustrates how the same facts have been treated differently across jurisdictions: registered in India, later contested; refused outright in the United States; accepted without substantive review in Canada. No settled international consensus on AI co-authorship has yet emerged from these divergent outcomes.

3.1.6 India: The Thaler Application

The Indian Copyright Office has separately been asked to consider Stephen Thaler's bid to register 'A Recent Entrance to Paradise', the same artwork refused registration in the United States in the Thaler litigation discussed in Section 3.1.1, this time on Thaler's claim that it was generated autonomously by his 'Creativity Machine' without any human author. Thaler filed the Indian application, Diary No. 9356/2022-CO/A, on 29 March 2022. It remained marked for 'hearing' for several years; on 9 April 2026, the Delhi High Court directed the Registrar of Copyrights to decide the application within eight weeks, and as of the most recently available records no decision had yet been reported. Its eventual outcome may set an important precedent for the copyrightability of AI-generated content in India.

3.2 Disclosure and Attribution Requirements

A separate, and equally important, issue concerns disclosure practice in the publishing sector, one that does not depend on the copyrightability analysis set out above. A growing number of publishers have adopted policies requiring authors to disclose which parts of a work were created using AI.

Leading academic and trade publishers have converged, in broad outline, on a common approach: generative AI tools cannot be credited as author or co-author, since authorship carries an accountability that only a human contributor can bear, but their use in preparing text, code, or images must ordinarily be disclosed, typically in the acknowledgements or methods section of the work[43]. This is consistent with the position taken by the Committee on Publication Ethics, whose guidance likewise excludes AI tools from authorship while requiring transparent disclosure of their use. For a publisher, that disclosure serves a purpose distinct from copyright compliance: it allows the publisher to identify, at the acquisition stage, which parts of a work may fall outside copyright protection under the standards discussed in Section 3.1, and to structure its representations, warranties, and indemnities accordingly, as discussed further in Chapter 6.2.

Disclosure alone does not guarantee copyright protection for AI-generated or AI-assisted work; that protection continues to depend on the legal standards set out in Section 3.1.

[42] *Li v. Liu, (2023) Jing 0491 Min Chu No. 11279 (Beijing Internet Court, 27 Nov. 2023).*

3.3 Implications for Publishers' Acquisition and Licensing Agreements

The analysis above yields three main drafting priorities for publishing agreements covering AI-generated works, discussed in more detail in Chapters 6.1 and 6.2.

(i) Publishers should ensure that authors disclose which generative AI tools were used in creating the content, together with the nature and extent of that use.

(ii) To support a good-faith claim of authorship, the agreement should include a warranty that the author exercised sufficient creative control over the final expression of the content.

(iii) The agreement should also include an indemnity clause allocating commercial risk in the event the work is later found ineligible for copyright protection.

Together, these measures let publishers and authors manage commercial risk while copyright law continues to evolve worldwide.

3.4 A Practical Record-Keeping Framework for Evidencing Human Authorship

The comparative practice discussed above shows that copyright offices increasingly look for contemporaneous evidence of what a human contributor actually did, rather than relying on what the applicant asserts after the fact. This section sets out practical steps that authors, publishers, and development teams can take to build the strongest possible record of human creative control.

Preserve the human-authored seed work separately. Save and timestamp any human input, such as a sketch, draft, photograph, or code skeleton, before AI involvement, and register it separately where it has standalone value. Suryast illustrates the point: Sahni's underlying photograph was itself a separate, independently registrable work, fixed before Raghav's style transfer was applied, and it would have survived even a full refusal of the composite image.

Document each human modification, with reasoning. For every change made to a raw AI output, record what was changed, why, and what alternative was considered or rejected, in the same file or commit as the change itself rather than in a retrospective summary written later. The Review Board in Suryast attributed the work's expressive choices to Raghav specifically because the record showed no modification by Sahni after generation; a contemporaneous edit log is the most direct way to avoid that outcome.

Log prompts, parameters, and tool settings. Record system and role messages, the model name and version, temperature and seed values, and any files provided as context, ideally captured automatically through the tool's API. A complete log demonstrates the full extent of creative direction supplied to the model.

Preserve the raw AI output separately from the edited version. Commit each raw generation to a dedicated location before any human editing begins, and maintain a clear differential between the raw and finished versions, so that the human's contribution is easy to identify and to credit.

Record model, version, and training data information. Capture the tool, model platform, version, and any vendor documentation on training data. Gaps in this information have counted against applicants before: the Review Board in Suryast noted that the Office did not know what preexisting works Raghav had been trained on, and treated that gap as a barrier to assessing originality.

Record selection, arrangement, and structural decisions. Document choices about organisation, ordering, hierarchy, and integration, and, where AI-generated candidates are chosen from a larger pool, log the candidates considered and the reasons for the final selection. Selection, coordination, and arrangement of this kind is recognised, including in *Zarya of the Dawn* and under the Office's own Part 2 Report discussed in Section 3.1.1, as a basis for protection even where individual AI-generated elements are disclaimed.

Disclose AI use accurately on the application. Identify content generated by AI that is more than de minimis, and claim only the human-authored contribution, consistent with the Office's March 2023 Registration Guidance discussed in Section 3.1.1. Accurate disclosure protects the registration against later cancellation of the kind seen in *Zarya of the Dawn*.

None of these steps guarantees registration. Where an AI tool alone determines the expressive elements of a work, with no human curatorial or directional input of the kind described above, current law in most jurisdictions will not protect the result, and no amount of record-keeping can substitute for that missing contribution. A disciplined contemporaneous record is what allows a genuine claim of human authorship to be evaluated on its merits.

Chapter IV: Research Integrity in the Age of AI



4.1 Defining Research Integrity and Research Misconduct

Research is the backbone of all professional pursuits. It is something which is common across all disciplines and professional pursuits. Whether it is for academia, or reading-intensive professional pursuits like legal services, or on the manufacturing side of things, research is something which simply cannot be dispensed with. As such, there is something very sacred about the integrity of research and since the inception of the written word, humanity has taken great care to ensure that research is error-free and logical, and most importantly, not fabricated.

With the advent of artificial intelligence in its current form, such as Claude, ChatGPT, and Perplexity, there has been a marked shift in how research is regarded. Research is now approached with a greater degree of scepticism than before, largely because AI has made the fabrication of research materials considerably easier. Given that the use of AI is not only widespread but is likely to grow substantially in the coming years, it is important to consider what research integrity and research misconduct mean in this context. Research integrity is what allows the scholarly record to be trusted, and in the age of AI, that trust can no longer rest on the assumption that a human alone produced, checked, and stands behind every part of it.

Research integrity frameworks, including those maintained by the Committee on Publication Ethics (COPE), have long drawn a line between honest error and research misconduct[44]. An honest error is a good-faith mistake, the kind ordinary scholarly processes can catch and correct. Misconduct is something else: it is conventionally understood to involve fabrication (inventing data or results), falsification (manipulating data, materials, or processes so the record no longer accurately reflects the research), or plagiarism (using someone else's ideas, methods, results, or words without giving proper credit).

Generative AI unsettles this distinction, because the whole framework rests on intent. Take an author who submits AI-generated citations without checking them, or AI-drafted text without disclosing it. In neither case did the author necessarily set out to deceive anyone, at least not in the way the traditional misconduct categories assume. And yet the record they produce is still wrong, or still improperly attributed, regardless of what they meant to do.

Thus, research integrity and research misconduct can be defined as below:

Research Integrity	Research Misconduct
The core tenet of research integrity is the commitment to conducting, recording, and reporting research honestly, accurately, and accountably. Given the adoption, usage and mass proliferation of AI in the realm of research, research integrity means checking rather than trusting: an AI-generated citation or a fluent AI-drafted paragraph is not verified just because it reads well, and the burden of confirming it is correct still falls on the person putting their name to it. Further, it means giving proper credit where credit is due. In this vein, it also means being	Research misconduct means intentionally distorting the research record. It is not an honest mistake. It covers three things: - Fabrication - Falsification - Plagiarism. The common thread is intent. If a researcher makes a mistake and fixes it, that is not misconduct. If they know something is wrong and publish it anyway, it is. AI makes this harder to apply. An AI tool can generate a fake citation. If a researcher

[44] https://publicationethics.org/news-opinion/research-misconduct

upfront about where and how AI contributed. Lastly, it also means accountability does not go away just because a tool did some of the work. For instance, a researcher who delegates a task to an AI has not delegated the responsibility for getting it right.	submits it without checking, they didn't "fabricate" anything themselves, but the fake citation is still in the record, under their name. Similarly, AI-written text can copy from other sources without credit.

4.2 Traditional Threats: Plagiarism, Data Fabrication, and Falsification

Before looking at what is new about AI-specific threats, it is worth stepping back to the baseline that already existed. Plagiarism, in the research-integrity sense, has never depended on copyright law. Rather, it is an ethical and institutional wrong in its own right.

Data fabrication and falsification are not new either. They existed long before generative AI, and they are still handled the same way institutions have always handled them: retraction, correction, or referral for discipline. That does not change just because AI happened to be involved in producing the fraudulent material.

For basic benchmarking and understanding the differentiation, please refer to the below basic examples:

Threat	Traditional	AI Based
Plagiarism	A doctoral candidate is writing the literature review section of her thesis. She finds a well-phrased passage in a published journal article that captures exactly the point she wants to make. Rather than paraphrasing it or setting it off as a quotation, she copies the passage nearly word-for-word into her own draft, changing only a few connecting words, and does not cite the original author. The ideas and language are not hers, but the thesis presents them as if they were. This is plagiarism in its classic form: a human being, consciously, took another human being's words and presented them as their own. **Example:** Major Ravi – *Karmayodha* (Kerala, 2025): A Kerala court held that the screenplay of the Malayalam film *Karmayodha* had been plagiarized	A researcher asks a generative AI tool to (Claude, ChatGPT, etc.) "write a paragraph summarizing the current literature on XYZ." AI produces a fluent paragraph and the same prima facie appears to be legitimate. However, the researcher being blissfully unaware, the paragraph closely mirrors, in structure and in several distinctive phrases, for instance a passage from a specific published paper the model was trained on, without attributing it. The researcher inserts this AI-generated paragraph directly into their manuscript, presents it as their own analysis, and does not disclose that AI was used to draft it. This is plagiarism without conscious intent: the writer did not set out to copy, but the result, unattributed use of another's words and ideas, is the same. **Example:**

	from another writer's script, and ordered compensation.[45]	Walmart employment litigation (U.S., 2026): In an employment dispute against Walmart, a lawyer uploaded Walmart's discovery responses into an AI tool and relied on the AI's analysis to identify alleged deficiencies. The judge found that the lawyer had effectively copied and pasted the AI-generated critique without exercising independent professional judgment, describing it as a "perilous shortcut." The court emphasized that AI may assist lawyers but cannot replace their duty to personally review and verify legal arguments.[46]
Fabrication	Fabrication is making up data or results that were never actually produced, and presenting them as if they were real. Nothing is altered from something true, rather, it is invented from nothing. **Example:** Diederik Stapel (Netherlands, 2011) Stapel, a social psychologist, was found to have fabricated data in dozens of published studies over roughly a decade in several cases, and it emerged he never ran the experiments at all, and simply invented the results he expected to find. An institutional investigation led to over 50 retractions and his resignation.[47]	AI-based fabrication works the same way, except the tool does the inventing instead of the person. A researcher asks an AI system for data, an image, or a result and gets back something that looks real but is not. It still ends up in a **Example:** Lawyer submitting AI-generated fake case citations to the Supreme Court of India (reported 2025): A lawyer filed a response containing non-existent cases generated with AI. The lawyer later admitted embarrassment over relying on AI without verification.[48]
Falsification	Falsification is different from fabrication. Here, the data is real as it actually came from somewhere. But it is been changed. Maybe a result got cropped, or an inconvenient data point got dropped, or numbers got rearranged, but just enough that the record no longer shows what actually	AI-based falsification is the same idea, just with a tool doing the altering instead of a person. A real image or a real dataset goes in, and AI is used to touch it up, for instance by smoothing over a result that didn't quite fit, erasing something inconvenient, or generating a

[45] https://www.barandbench.com/news/kerala-court-orders-filmmaker-major-ravi-to-pay-30-lakh-compensation-for-plagiarising-screenplay-of-karmayodha
[46] https://www.reuters.com/legal/legalindustry/lawyers-use-ai-was-perilous-shortcut-walmart-case-us-judge-says-2026-04-15
[47] https://www.theguardian.com/science/2012/sep/13/scientific-research-fraud-bad-practice
[48] https://economictimes.indiatimes.com/tech/artificial-intelligence/never-been-more-embarrassed-lawyer-cites-fake-cases-in-ai-aided-response-to-sc/articleshow/125866590.cms

	happened. Nothing was made up from nothing. Something true was altered until it stopped being true. **Example:** Haruko Obokata incident: She was a Japanese stem-cell researcher at RIKEN who claimed in 2014 that she had discovered a simple method, called STAP cells, to convert ordinary cells into pluripotent stem cells. The findings were published in *Nature* but were later questioned after scientists found problems with the reported data, including manipulated and duplicated images. A RIKEN investigation concluded that Obokata had committed research misconduct through falsification of research data. The papers were retracted, she resigned from RIKEN, and the case became a major example of the importance of research integrity.[49]	slightly different version of the same image so it can be reused without looking like a repeat. The starting material was genuine. What comes out is not. **Example:** *Hypothetical:* A scientist conducts a real microscopy experiment and obtains genuine images. Before publishing, they use an AI tool to remove unwanted features or enhance results so that the findings appear stronger than they actually are. The experiment was real, but the evidence was altered. This is **falsification**, not fabrication, because existing data was manipulated rather than invented.

4.3 Plagiarism in the AI Era

4.3.1 Traditional Plagiarism and Copyright Overlap

Plagiarism and copyright infringement are often treated as interchangeable. However, it is not so entirely, although there is notable overlap and there are grey areas. The very basic differentiation however, is that plagiarism, at its core, is a question of credit: did the author acknowledge where the words or ideas came from, or allow the reader to assume they were original. To simplify: all copyright infringement comprises of plagiarism, however, all plagiarism may not amount to infringement.

Copyright infringement operates on an entirely different basis. It is a legal wrong. It requires unauthorised copying of something the law protects, i.e. specific expression, not the underlying idea. A court determines infringement using statutory tests: whether the original work was protectable, whether the copying was substantial, whether an exception such as fair use or fair dealing applies. None of this turns on whether credit was given. For instance, in works such as research papers. The above reflects a rather traditional view of plagiarism as against copyright. The advent of generative AI does not change that underlying definition, as the question of whether credit was given remains exactly what it was. What has changed is how easily the underlying conduct can now be concealed. A researcher can upload existing text of another author's work, or their own earlier publication, into a tool such as Claude or ChatGPT and request a complete rewrite, producing a version that reads as freshly composed while remaining, in substance, the same work. The wrong is unchanged. What AI has altered is the ease of committing it invisibly.

[49] https://www.riken.jp/en/news_pubs/research_news/pr/2014/20140314_1/index.html

4.3.2 AI-Assisted Plagiarism and Undisclosed AI-Generated Content

The central new risk in this category follows directly from how large language models are built. These models are trained on a vast corpus of text that includes copyrighted expression belonging to countless other authors.

The risk arises when a researcher takes the model's output, such as a paragraph, a literature review, an entire section of analysis, and submits it as original scholarship, without disclosing that any part of it was machine-generated. Stated this way, the problem is simple. Detecting it, in practice, is not.

The difficulty is not incidental as it follows from the mechanics of the tools available. Similarity-detection software such as Turnitin operate by comparing submitted text against a database of existing sources and flagging passages of direct verbal overlap. This approach was designed to catch a particular, older form of misconduct: a student copying several paragraphs from a published article and altering a handful of words. It was not designed to catch a model that has absorbed the phrasing and argumentative patterns of thousands of sources and produces a passage that resembles all of them in substance while matching none of them closely enough in wording to register as a match. The resulting text can appear entirely original at the sentence level, and yet not be the work of the person whose name appears on it.

This risk is not merely theoretical in the current environment of widespread AI use; such practices are understood to occur at scale, and the resulting consequences are not confined to industry and business settings. One significant and comparatively under-discussed consequence is the following:

- Lowering intellectual standards amongst students; and

- Building dependence on AI.

Instances illustrating this dynamic have surfaced in a number of jurisdictions; two examples from India, discussed below, are illustrative of the broader pattern:

- O.P. Jindal Global University (2024): An LLM student challenged the university after his examination answer sheet was declared **AI-generated**, leading to litigation before the Punjab and Haryana High Court.[50]

- Lucknow University PhD screening (2026): a university review found that a majority of that year's plagiarism flags were attributable to AI or similarity concerns, illustrating the scale of the detection challenge described above.[51]

The underlying difficulty, in every jurisdiction, is the same. The safeguard that academic and publishing institutions have relied upon for decades was built to answer one question: do these words resemble words that already exist elsewhere? It was never built to answer a different, more fundamental question, i.e. whether the words originated with the author at all. It is this second question which this new category of misconduct places beyond the reach of the existing tools.

4.3.3 Paraphrasing Tools, AI Rewriting, and Attribution Challenges

A related risk, and in some respects a more deliberate one, involves the use of AI-powered paraphrasing tools to rewrite existing text. For instance, sometimes text belonging to another

[50] https://indianexpress.com/article/cities/chandigarh/ai-punjab-haryana-hc-jindal-student-failed-9653575/?utm_source=chatgpt.com
[51] https://timesofindia.indiatimes.com/city/lucknow/academic-integrity-alert-lu-flags-plagiarism-in-95-of-phd-theses-submitted-this-year/articleshow/129665165.cms?utm_source=chatgpt.com

author, sometimes the researcher's own earlier work, for the specific purpose of evading similarity detection while leaving the underlying substance untouched.

The practice is not new in itself; in the context of search-engine optimisation and content farming, it has long been known as content spinning, and tools built for exactly this purpose have existed since well before generative AI entered academic use. What has changed is the sophistication of the output. Earlier spinning tools produced text that was often clumsy or grammatically strained, and readable as manipulated on inspection. Modern AI paraphrasers produce fluent, natural-sounding prose that gives no outward sign of having been generated from a source text at all. Simply put, to most untrained human eyes and majority of anti-plagiarism software, it is extremely difficult (and in a fair few cases, nearly impossible) to discern fully original works from AI assisted/ generated work which has been 'plagiarised' (in the modern sense) from pre-existing works.

This raises the attribution question in its most acute form. A passage run through such a tool may share no meaningful string of words with its source, and will accordingly return a low or negligible score on a conventional similarity check. The tool has not helped the author say something in their own words; it has helped them make somebody else's words look like their own.

The common thread running through this chapter and the one preceding it is that similarity detection was built to answer a single question: does this text resemble something that already exists. Undisclosed AI generation defeats that question because the output resembles no single source closely enough to be flagged. Paraphrasing tools defeat it by design, taking a text that would otherwise be flagged and deliberately reducing that resemblance while preserving everything a reader would recognise as the substance of the work.

Examples of few such tools are listed below:

- QuillBot

- WordAi

- Spinbot

- Paraphraser.io

- Wordtune

4.3.4 Similarity Detection Tools and Their Limitations

In modern publishing, a single number or percentage can carry paramount importance. A manuscript or draft is submitted, the integrity check runs overnight, and by morning a figure is waiting: 27%, or 4%, alongside further reported figures such as a 70% likelihood of AI generation. The central difficulty in this era of fluent AI writing and 'humanised' outputs is that a degree of uncertainty is likely to remain over the results produced by current detection tools.

Similarity checkers and AI detectors sit side by side in the same interfaces, but they are not equivalent instruments. A similarity checker makes a **falsifiable** claim. It points at something: *these forty-one words also appear, in this order, in that 2019 article*. An editor can open both documents. The author can respond. If the tool is wrong, the error is visible in seconds.

Whereas, an AI detector points at nothing. Its output is a probability derived from statistical features of the text, with no source document to inspect and nothing the accused can examine. An author facing a similarity report can say "look at page 14, I cited that." An author facing a

detection score can only say "I didn't." A process built on evidence the subject cannot contest is not a fair process, whatever its outcome.

Text overlap is a symptom, not a finding. Legitimate overlap comes from standard methods descriptions, reference lists, the author's own preprint, and quoted material. This last point is acute in Indian legal scholarship: every thesis on the right to life quotes Article 21 and *Maneka Gandhi v. Union of India*, often the same passages, because those are the passages that matter. A rigorous dissertation may score 22% consisting almost entirely of correctly attributed quotation. Thus, especially in the legal industry, such plagiarism detection may be in fact unfair to the writers/ authors whose work get flagged as plagiarised. After all, *precedents* is an integral and unignorable part of the legal ecosystem.

Given this subjectivity and the resulting uncertainty over AI-detection results, there is a clear need to develop more sophisticated detection tools. As generative AI becomes further embedded in academic and publishing workflows, detection methods will similarly need to evolve to address AI-assisted plagiarism effectively.

The central difficulty lies in the continued failure of AI-detection tools to reliably identify AI-generated content.

It is evident that the underlying problem is neither AI itself nor its proliferation. Rather, the difficulty lies in the inadequacy of current plagiarism-detection tools to reliably identify AI-based plagiarism.

This is a problem which makers of AI have also tried to address, most notably, OpenAI withdrew its AI Text Classifier in July 2023, citing low accuracy.[52]

The priority, therefore, is to refine detection protocols and tools in order to reduce their error rates.

4.4 New AI-Enabled Threats

4.4.1 AI-Generated ("Hallucinated") Citations and References

Of all the ways generative AI has disturbed the machinery of scholarly publishing, the fabricated citation is the easiest to demonstrate. Ask a large language model for authority on almost any proposition and it will supply one: authors, title, journal, volume, page range, year, often a DOI that looks entirely conventional. The formatting will be immaculate. The source will not exist.

It is ironic that it is the legal industry in many countries which has seen this problem rear its ugly head. For instance:

- In *Mata v. Avianca, Inc.*, before the Southern District of New York in 2023, a lawyer submitted a brief citing several decisions that no court had ever handed down! This was widely reported and the lawyer who submitted the fake citations was fined.[53]

- Delhi High Court (2025) – similar issue happened before the Hon'ble Delhi High Court, wherein a lawyer cited authorities which did not exist and the petition was withdrawn.[54]

[52] https://www.thehindu.com/sci-tech/technology/openai-pulls-back-ai-generated-text-detector-cites-low-rate-of-accuracy/article67126812.ece
[53] https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/
[54] https://economictimes.indiatimes.com/tech/artificial-intelligence/bogus-ai-case-citations-in-delhi-hc-raise-red-flags/articleshow/124239436.cms?utm_source=chatgpt.com&from=mdr

The scale of the issue is exacerbated by the fact that reportedly, the rate of fabrication by certain AI modes is quite high. For instance, a 2023 study in *Cureus* examined references generated by ChatGPT for short medical review articles and found roughly 47 per cent entirely fabricated, a further 46 per cent authentic but inaccurately reported, and only about 7 per cent both real and correctly cited.[55]

4.4.2 Fabricated Data and Synthetic Datasets

Generative AI's ability to produce plausible-looking structured data raises the risk of wholly fabricated datasets or results tables submitted as genuine empirical findings, and the same is a more sophisticated variant of traditional data fabrication.

Synthetic data refers to records generated by an AI model to mimic the statistical shape of real data. This may have positive applications, such as use by Hospitals to share patient-like records without exposing patients; statistical agencies release synthetic microdata so researchers can work without breaching confidentiality, etc.

In contrast, *fabricated data* is essentially invention passed off as observation. What generative AI has done is collapse the technical distance between the two. The same tool that protects privacy in one hand produces fraud in the other.

One notable well reported example of this is the paper released by *JAMA Ophthalmology* in 2023, wherein using GPT-4, they generated a fake dataset of patients undergoing two different corneal surgeries and instructed it to produce numbers favouring one procedure. The result was startlingly authentic looking.[56]

4.4.3 Paper Mills and AI-Assisted Mass Manuscript Generation

A paper mill is not a sloppy author or a careless lab. It is a business. Somebody writes a manuscript, often around fabricated or recycled data, and then sells slots on the author byline to academics who need a publication. Prices are advertised, positions are ranked (first author costs more than fourth), and delivery is promised within a fixed window.

For years, this looked like a nuisance at the margins of scholarly publishing. It no longer does. Between 2023 and 2024, Wiley retracted more than 8,000 papers linked to systematic manipulation across journals it had acquired from Hindawi, shut down four of those journals outright, and eventually retired the Hindawi imprint altogether.[57]

Paper Mills are hardly new or novel, they have existed for years. The problem is that with the advent and mass proliferations and use of AI, the dynamics of paper mills have changed significantly. The mills existed before ChatGPT. What large language models altered was throughput. Producing a plausible-sounding introduction, a literature review with real-looking citations, and a discussion section used to be the bottleneck; now it is close to free.

4.4.4 Undisclosed AI Authorship/Co-Authorship

Authorship in scholarly publishing has never been merely decorative. It is a claim of responsibility. It is an assertion that identifiable human beings conceived the work, stand behind its content, and can be held to account if it turns out to be wrong. Generative AI has strained that claim from two opposite directions at once.

[55] https://www.cureus.com/articles/158289-high-rates-of-fabricated-and-inaccurate-references-in-chatgpt-generated-medical-content#!/
[56] https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2804364
[57] https://retractionwatch.com/2023/12/19/hindawi-reveals-process-for-retracting-more-than-8000-paper-mill-articles/

The prevailing position among major journal publishers and international bodies is that an AI tool cannot be listed as an author, on the reasoning that authorship carries accountability obligations.

This is not a simple legal problem. This is a grave matter concerning ethics as well. It is tempting to treat non-disclosure as a technicality. It is not. The ICMJE criteria for authorship require four things: a substantial contribution to the work, involvement in drafting or revising it, approval of the final version, and agreement to be accountable for it.[58] A language model can plausibly do the first two. It cannot do the last two, and it is the last two that make authorship a moral position rather than a description of labour. Concealing machine involvement does not merely misstate who typed the words. It obscures where responsibility sits when something goes wrong. There is also a straightforward question of honesty toward readers.

Besides the above noted ethical and moral implications, another question which is unavoidable is that of attribution of AI as authors for the purpose of Copyright. This is something which has had notable discussions in India in the past few years. Section 2(d) of the Copyright Act, 1957 defines "author" work-type by work-type: the artist for an artistic work, the composer for a musical work, the person taking a photograph; and then, in clause (vi), inserted by the 1994 amendment, addresses computer-generated works directly: for any literary, dramatic, musical or artistic work that is computer-generated, the author is *the person who causes the work to be created*.

[58] https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html



Chapter V:

Regulatory Frameworks and Sector Impact in India

India's engagement with AI and copyright is taking shape through four largely independent tracks: the pre-digital text of the Copyright Act, 1957, which remains the primary source of substantive rights and defences; the intermediary liability provisions of the Information Technology Act, 2000, substantially amended to address synthetically generated content; the innovation-oriented policy architecture built around the IndiaAI Mission, within which intellectual property questions are now addressed directly by the Department for Promotion of Industry and Internal Trade (DPIIT); and the academic-integrity regulations issued by the University Grants Commission (UGC) and the All India Council for Technical Education (AICTE). Each track carries direct consequences for publishers operating in or licensing into the Indian market, whether as owners of assets scraped for AI training, as intermediaries hosting AI-assisted content, or as institutions responsible for the integrity of the scholarly record.

Taken together, these four tracks reflect active and ongoing engagement by Indian policymakers, courts, and regulators with the questions AI raises for copyright, rather than an absence of attention to the issue. The IndiaAI Mission's investment in AI infrastructure, DPIIT's stakeholder consultations on generative AI and copyright, and the Delhi High Court's substantive engagement with these questions in 2026 together illustrate a legal and policy landscape that is actively evolving, even as the framework continues to take shape.

5.1 Indian Legal and Policy Framework

5.1.1 The Copyright Act, 1957 and Copyright Rules

The Copyright Act, 1957 remains the governing statute for copyright protection in India. It has been amended on several occasions since its enactment, most notably in 1994, in 1999 to align with the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), and in 2012 to strengthen protection in the digital environment. None of these amendments anticipated machine learning systems trained on large corpora of text, images, or audio, and the statute continues to assume discrete human authorship and comparatively limited-scale reproduction.

Section 14 confers on the copyright owner the exclusive right to reproduce the work in any material form, including electronic storage, together with rights of adaptation and communication to the public. Ingesting copyrighted works into a training corpus and storing them to train a model may fall within this reproduction right. Unauthorised exercise of these rights constitutes infringement under Section 51, attracting civil remedies and, in certain circumstances, criminal liability.

The principal statutory defence available to an AI developer is Section 52, India's fair dealing regime. Unlike the open-ended fair use doctrine under US law, Indian fair dealing is a closed list of enumerated purposes. Section 52(1)(a)(i) exempts fair dealing with any work, not being a computer programme, for private or personal use, including research. Whether large-scale commercial processing of copyrighted works to train a large language model qualifies as "private or personal use, including research" has become the central doctrinal question in India and received its first significant judicial consideration in 2026. India has no standalone text and data mining exception comparable to those adopted in jurisdictions such as the European Union, Japan, or Singapore.

This case law predates the generative AI debate but continues to frame it. In *R.G. Anand v. Delux Films*, the Supreme Court held that infringement requires substantial copying of protected expression rather than merely an underlying idea. In *Eastern Book Co. v. D.B. Modak*, the Court rejected the "sweat of the brow" doctrine and instead required skill and

judgment involving a modicum of creativity, a standard formulated with human authorship in mind. The Delhi High Court's decision in *University of Oxford v. Rameshwari Photocopy Services* and the Kerala High Court's decision in *Civic Chandran v. Ammini Amma* likewise illustrate the narrow, purpose-specific interpretation Indian courts generally adopt when applying Section 52. Indian copyright law does not recognise a freestanding doctrine of transformative use, making it more difficult to argue that AI training is automatically exempt because of its technical or non-expressive character.

These questions moved from theory to active litigation in *ANI Media Pvt. Ltd. v. OpenAI Inc. and Others*, filed before the Delhi High Court in November 2024, in which ANI alleged that OpenAI's use of its news reports to train ChatGPT infringed its copyright under Section 14. On 24 July 2026, after 32 hearings, Justice Amit Bansal dismissed ANI's application for an interim injunction, holding on a prima facie basis that OpenAI's storage of ANI's works for training large language models fell within the fair dealing exception for "private or personal use, including research" under Section 52(1)(a) and therefore did not amount to infringement under Section 51.[59] The Court further held, on a prima facie basis, that ANI had not established substantial reproduction of its works in ChatGPT's outputs and confirmed that the Delhi High Court had territorial jurisdiction despite OpenAI's servers being located outside India.

The Court expressly confined these findings to the interim stage, and the suit remains pending for final adjudication. Nevertheless, the order represents the first substantive judicial consideration of AI training under India's fair dealing provisions.[60]

In parallel, DPIIT constituted an expert committee in April 2025 to examine the relationship between generative AI and copyright, a development discussed further in Section 5.1.3.

On the output side, the Act does not expressly address works generated wholly or substantially by an AI system. Section 2(d)(vi) defines the "author" of a computer-generated work as the person who causes the work to be created, language borrowed from section 9(3) of the UK's Copyright, Designs and Patents Act 1988 and discussed in Chapter 3. The Copyright Office's handling of the RAGHAV application, which listed an AI system as a co-author before subsequently clarifying its position, is also discussed in that chapter.

5.1.2 The IT Act, 2000 and Intermediary Liability

The second pillar operates through the intermediary liability regime under the Information Technology Act, 2000. Section 79 gives intermediaries, defined broadly under Section 2(1)(w), a conditional safe harbour, provided they comply with the prescribed due diligence requirements and do not initiate the transmission, select the receiver, or modify the information contained in the transmission. Rule 3(1)(b) of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 requires intermediaries to make reasonable efforts to prevent the hosting of content that infringes copyright or other proprietary rights. Failure to comply with these obligations may result in the loss of safe harbour, exposing the intermediary to liability under the applicable law.

Before introducing binding regulation, the Ministry of Electronics and Information Technology (MeitY) addressed generative AI through a series of advisories, beginning with its 7 November 2023 advisory on deepfakes, followed by the advisory of 1 March 2024, which was revised on 15 March 2024 to clarify expectations regarding generative AI, misinformation, and under-

[59] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024.*
[60] *Id.*

tested AI models. These advisory measures were subsequently given a statutory basis through the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified on 10 February 2026 (G.S.R. 120(E)) and brought into force on 20 February 2026.[61] The amendments introduce "synthetically generated information" as a defined category under new Rule 2(1)(wa) and impose obligations relating to the disclosure, labelling, provenance, and verification of such content.

Under Rule 4(1A), before allowing a user to display, upload, or publish information, significant social media intermediaries must obtain a user declaration on whether the content is synthetically generated, and must deploy reasonable and appropriate technical measures to verify that declaration before publication. Confirmed synthetic content must then carry a clear and prominent label. Separately, under new Rule 3(3)(b), an intermediary offering tools that enable the creation of synthetic content must not permit the modification, suppression, or removal of the labels, permanent metadata, or provenance identifiers embedded in that content.[62] The amendments also shorten takedown timelines for the most harmful categories of content: from 24 hours to 2 hours for complaints involving nudity, sexual content, or impersonation under Rule 3(2)(b), and from 36 hours to 3 hours for removal upon actual knowledge under Rule 3(1)(d). These timelines apply to intermediaries operating in or offering services to users in India.

The 2026 Rules assume that AI-enabled platforms fall within the statutory definition of an "intermediary" without distinguishing between platforms that merely host AI-generated content and those that develop or operate generative AI models. Whether that distinction should affect the availability of safe harbour remains an open legal question. For publishers hosting third-party or AI-assisted content, the practical effect is an expanded due diligence obligation in addition to the pre-existing requirements under Rule 3(1)(b). A separate reform proposal, the Digital India Bill, first previewed in 2023, contemplated a differentiated regulatory framework based on platform risk, but it has not yet been enacted. Accordingly, intermediary liability in India continues to be governed principally by Section 79 of the IT Act as supplemented by the 2021 Rules and the 2026 amendments.

5.1.3 India AI Mission and IP Policy

India's principal AI industrial policy initiative is the IndiaAI Mission, approved by the Union Cabinet in March 2024 with an outlay of ₹10,372 crore.[63] The Mission supports the development of compute infrastructure, datasets, foundation models, AI applications, skilling, and "Safe & Trusted AI."[64] It builds on NITI Aayog's National Strategy for Artificial Intelligence (2018)[65] and operates alongside the National IPR Policy of 12 May 2016.[66] While the Mission promotes AI development and adoption, it does not establish rules governing copyright. Issues relating to copyright and generative AI have instead been examined through a separate policy process led by the Department for Promotion of Industry and Internal Trade (DPIIT).

---

[61] *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.*
[62] *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as amended, Rules 3(3)(b) and 4(1A).*
[63] *https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2012375&reg=48&lang=2*
[64] *https://indiaai.gov.in (accessed 4 August 2026).*
[65] *NITI Aayog, National Strategy for Artificial Intelligence: #AIforAll, June 2018: https://www.niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf (accessed 4 August 2026).*
[66] *Government of India, National Intellectual Property Rights Policy, Department for Promotion of Industry and Internal Trade, 12 May 2016.*

In April 2025, DPIIT constituted an eight-member expert committee to examine the relationship between generative AI and copyright law.[67] In December 2025, the Committee published Working Paper on Generative AI and Copyright – Part I: "One Nation, One License, One Payment."[68] The Working Paper is limited to the use of copyrighted works as inputs for AI training and proposes a hybrid licensing framework. Under the proposal, AI developers would receive a mandatory blanket licence to use lawfully accessed copyrighted works for training, without an opt-out mechanism, while copyright owners would receive statutory remuneration linked to the commercialisation of AI systems.[69] To administer this framework, the paper proposes the establishment of a Copyright Royalties Collective for AI Training (CRCAT) and the introduction of an AI Training Data Disclosure Form.[70]

The subsequent public consultation elicited a range of stakeholder responses.[71] Some rights holders and creator organisations supported the proposal on the basis that it would provide guaranteed statutory remuneration. Other stakeholders, including certain technology developers and open-source advocates, expressed concerns regarding the absence of an opt-out mechanism, the lack of protection for personality rights, and the potential incompatibility of mandatory licensing with open-source and share-alike licensing models.[72]

The Working Paper remains a policy proposal and does not have legal effect. Its recommendations would require legislative amendment or subordinate legislation before implementation.[73] Accordingly, both the publication of the Working Paper and the consultation process reflect an ongoing and iterative policy development process rather than a settled regulatory position. Nevertheless, the proposal indicates consideration of a possible shift from an unlicensed text and data mining model towards a compensated-access framework resembling extended collective licensing, discussed further in Chapter 6.

Among the consultation responses, the FICCI Publishing Committee submitted comments on behalf of publishers representing the trade, academic, educational, and regional-language sectors. The Committee argued that voluntary licensing arrangements are already functioning in practice and should not be displaced by an untested mandatory licensing regime, particularly given the publishing sector's reliance on negotiated licensing that reflects differences in genre, language, and the commercial value of individual works rather than a single administratively determined rate.

The submission also questioned whether a centralised royalty collective could effectively identify and compensate rightsholders beyond existing copyright societies, particularly those operating in regional languages. It further observed that a system based on deferred statutory remuneration could permit copyrighted works to be used for AI training before compensation is received. At the same time, the Committee supported the introduction of a mandatory AI developer registry and dataset disclosure requirements as measures to improve transparency and accountability. It recommended strengthening voluntary licensing arrangements and

---

[67] *https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2200741&lang=2&reg=48 (accessed 4 August 2026).*

[68] *Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright – Part I: One Nation One License One Payment, December 2025: https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf (accessed 27 July 2026).*

[69] *Id.*

[70] *Id.*

[71] *https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2200741&lang=2&reg=48*

[72] *Software Freedom Law Center, India, Comments on DPIIT Working Paper on Generative AI and Copyright, Part 1, SFLC.in, 9 February 2026: https://sflc.in/comments-on-dpiit-working-paper-on-generative-ai-and-copyright-part-1/*

[73] *Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright – Part I: One Nation One License One Payment, December 2025: https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf*

enforcement mechanisms in preference to a mandatory blanket licensing model. These issues are examined further in the collective licensing analysis in Section 6.9 and the broader policy assessment in Chapter 7.

5.1.4 UGC and AICTE: Academic Integrity

The fourth regulatory track concerns academic integrity and research misconduct. The University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations, 2018 require higher educational institutions to establish mechanisms for plagiarism detection, screen theses, dissertations, and other academic work for plagiarism, and classify plagiarism into graded levels with corresponding academic penalties.[74] The Regulations, however, were enacted before the emergence of generative artificial intelligence and regulate plagiarism primarily through standards of originality and textual similarity. They contain no provisions addressing AI-generated authorship, disclosure of AI assistance, or the permissible use of generative AI tools in academic writing.[75]

In the absence of AI-specific national regulations, universities have increasingly developed their own institutional policies governing the use of generative AI in teaching, assessment, and research. Consequently, the treatment of AI-assisted academic work presently depends largely upon institutional academic integrity policies rather than a uniform regulatory framework issued by the UGC. Although some universities have reportedly applied existing plagiarism rules or institutional regulations to AI-generated submissions, these practices are institution-specific and do not derive from any UGC notification.

AICTE has adopted a different policy orientation by encouraging AI adoption and capacity building in technical education. Through its policy initiatives, AICTE has promoted the integration of artificial intelligence into higher education by encouraging AI-ready curricula, faculty development, institutional AI strategies, and industry collaboration.[76] As part of these initiatives, AICTE designated 2025 as the "Year of Artificial Intelligence" and encouraged AICTE-approved institutions to prepare institution-specific AI implementation plans to strengthen AI-enabled teaching, research, and governance.[77]

AICTE has also partnered with OpenAI under the OpenAI Learning Accelerator initiative. According to OpenAI, the programme includes the provision of approximately 500,000 ChatGPT licences for students and educators across India, educator training programmes, support for government schools and AICTE-approved institutions, and a US$500,000 research collaboration with IIT Madras. Contemporary reports further indicate that approximately 150,000 ChatGPT Go licences are specifically allocated to AICTE-affiliated institutions as part of the broader initiative.[78] While these initiatives demonstrate governmental and institutional support for AI adoption in education, they do not establish binding legal standards governing academic integrity or AI-generated content.

[74] University Grants Commission, The University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations, 2018: https://www.ugc.gov.in/e-book/UGC_Regulation/files/basic-html/page512.html
[75] Ibid.
[76] All India Council for Technical Education, AICTE Inputs on Artificial Intelligence (Comprehensive): https://www.aicte.gov.in/downloads/initiatives/AICTE%20inputs%20on%20Artificial%20Intelligence%20-comprehensive.pdf
[77] All India Council for Technical Education, Year of Artificial Intelligence 2025. https://www.aicte.gov.in/downloads/initiatives/AICTE%20inputs%20on%20Artificial%20Intelligence%20-comprehensive.pdf
[78] The Times of India, 'OpenAI announces Learning Accelerator program in India: 5 lakh ChatGPT licences for Indian students and educators': https://timesofindia.indiatimes.com/technology/tech-news/openai-announces-learning-accelerator-program-in-india-5-lakh-chatgpt-licences-for-indian-students-and-educators/articleshow/123542877.cms

Accordingly, the present regulatory landscape remains bifurcated. Academic integrity continues to be governed principally by the UGC's pre-generative AI plagiarism regulations, whereas national education policy increasingly encourages responsible AI adoption through institutional partnerships and capacity-building initiatives. The absence of nationally applicable rules on AI-generated authorship or mandatory disclosure means that questions relating to AI-assisted academic work continue to be determined primarily by individual universities, publishers, editors, and examining bodies.

Overall, India's regulatory approach reflects a gradual transition rather than a regulatory vacuum. While the UGC's existing plagiarism framework continues to safeguard academic integrity, AICTE's initiatives demonstrate a parallel commitment to promoting the responsible adoption of artificial intelligence in higher education. Together, these developments indicate an emerging policy direction that seeks to balance academic honesty with technological innovation.

Chapter VI:
Contracts, Liability and Dispute Resolution



6.1 The Contractual Problem

Publishing agreements were developed around a relatively stable assumption: the manuscript delivered by an author would reflect human intellectual effort, while the publisher would undertake editing, production, distribution and marketing through human-led processes. Generative artificial intelligence unsettles that assumption. A manuscript may now contain text, images, translations, summaries or research generated with varying degrees of machine assistance. The publisher may also use AI tools for editorial assessment, metadata, marketing, discoverability, translation or audiobook production. Each use creates a distinct combination of copyright, confidentiality, data protection, accuracy and reputational risk.

Traditional originality and non-infringement warranties remain relevant, although they rarely answer the operational questions that arise before publication. They may not identify the tool selected by the author, the material uploaded to that tool, the extent of machine-generated content, the review carried out by the author, or the rights retained by the tool provider. A carefully designed agreement should therefore regulate process as well as outcome.

The Indian legal position requires particular care. The Copyright Act, 1957 recognises the person who causes a computer-generated literary, dramatic, musical or artistic work to be created as its author. That provision predates modern generative AI and has not yet produced a settled judicial test for the degree of human direction required for copyright protection in AI-assisted works. The Act continues to reserve to copyright owners the exclusive rights set out in section 14 and regulates ownership, assignment and licensing through sections 17 to 19 and section 30.[79] Contract drafting should accommodate this uncertainty rather than state that every AI-assisted output is either protected or unprotected.

A useful contract begins by separating five transactions that are often grouped under the expression “AI use”:

(a) an author using a tool while preparing a manuscript;

(b) a publisher processing a manuscript through an AI-enabled workflow;

(c) a publisher licensing its catalogue for model training or retrieval;

(d) an AI developer supplying a model or service; and

(e) an end user producing an output.

Rights and responsibilities should be allocated to the party controlling the relevant act, information and safeguard.

6.2 Classification of AI Use in Publishing

A prohibition covering every use of AI is difficult to administer because spell-checking, grammar tools, transcription and search functions may already incorporate machine-learning features. A more workable approach classifies use by its effect on authorship, confidentiality and publication risk.

79 Copyright Act, 1957, ss 2(d)(vi), 14, 17 to 19 and 30, Copyright Office, Government of India, https://copyright.gov.in/documents/copyright_act_1957.pdf (accessed 27 July 2026).

Category	Illustrative use	Contractual treatment
Routine assistance	Spelling, grammar, formatting, transcription and non-substantive copy-editing.	Permitted without separate disclosure, subject to confidentiality and security requirements.
Material assistance	Research summaries, translation, structural suggestions, substantive rewriting, image generation, metadata and marketing copy.	Permitted subject to disclosure, verification and any approval requirements stated in the agreement.
Restricted use	Generation of a substantial part of the work, imitation of an identifiable author, uploading confidential material, fabricated citations, or use of tools whose terms permit unsuitable retention or reuse.	Prohibited unless the publisher gives prior written approval and appropriate safeguards are documented.

The classification should be applied by reference to effect rather than the marketing description of a product. A tool may describe itself as an editor while generating substantial new expression. Conversely, a technically AI-enabled feature may make only mechanical corrections. The agreement should therefore define **"Material AI Use"** by reference to whether the tool generates, translates, materially rewrites or determines substantive content.

Disclosure should operate as a compliance mechanism. Treating every disclosure as an admission of breach will encourage incomplete reporting. The author should remain responsible for review and verification, while the publisher should retain a proportionate right to request supporting information, further editing or removal of material that presents a documented legal or editorial concern.

6.3 Model Clauses for Author and Publishing Agreements

6.3.1 Definitions

Model clause: Definitions

"Generative AI System" means any software, model or service that generates or materially transforms text, images, audio, video, code, translations or other expressive content in response to prompts, examples or other inputs.

"Material AI Use" means use of a Generative AI System to generate, translate, summarise, substantively rewrite or determine any material element of the Work. It excludes use confined to spelling, grammar, formatting, transcription and other non-substantive mechanical assistance.

"AI-Generated Material" means content generated or materially transformed through Material AI Use and incorporated, in whole or in part, into the Work.

6.3.2 Author disclosure and review

Model clause: Author use and disclosure

The Author shall disclose to the Publisher any Material AI Use in the preparation of the Work. The disclosure shall identify the Generative AI System used, describe the purpose and extent of use, and identify the portions of the Work materially affected.

The Author shall exercise independent intellectual judgement and sufficient creative control over the Work. Before delivery, the Author shall review and verify all AI-Generated Material, including factual statements, quotations, citations, attributions and references.

The Author shall not knowingly include fabricated citations, quotations, authorities or factual claims. The Author shall not upload to a Generative AI System any confidential information, personal data, unpublished third-party material or content that the Author is not authorised to disclose.

Disclosure of Material AI Use shall not by itself constitute a breach. Where the disclosed use creates a material legal, editorial or reputational concern, the Publisher may reasonably require supporting information, further review, amendment or removal of the affected material.

The phrase "sufficient creative control" should be read with the authorship framework developed elsewhere in this book. Within the agreement, it performs a practical function. The author confirms active selection, judgement and revision without offering an absolute warranty on a legal question that may remain unsettled.

6.3.3 Publisher use of the manuscript

The agreement should regulate the publisher's own use of AI. In April 2026, the Authors Guild published model clauses addressing the uploading of manuscripts and personal information to consumer-facing AI systems and substantive AI editing. Its stated position favours written author permission and safeguards against training use.[80] That approach reflects a wider contractual concern: the author may have granted publication rights without consenting to the manuscript becoming training material or being exposed through a third-party service.

Model clause: Publisher use of AI

The Publisher shall not upload the Work, unpublished drafts or the Author's personal information to a consumer-facing Generative AI System without the Author's prior written consent.

Where the Author gives consent, the Publisher shall use reasonable contractual and technical safeguards to prevent the Work and personal information from being retained for general model training or disclosed to unauthorised persons.

The Publisher shall not use a Generative AI System to make substantive textual or artistic changes to the Work without the Author's prior written approval. Routine spelling, grammar, formatting and production checks may be undertaken through approved tools that satisfy the Publisher's confidentiality and security requirements.

Any permitted use shall remain subject to the Publisher's obligations concerning confidentiality, data protection, attribution and editorial responsibility.

80 https://authorsguild.org/news/use-of-ai-in-publishing-and-new-model-contract-clause/

6.3.4 Reservation and grant of AI rights

Broad grants covering all media and technologies may generate disputes over whether training, fine-tuning, retrieval or synthetic data uses were contemplated. AI rights should be identified expressly. The Authors Guild's model clauses similarly recommend express negotiation of AI-related rights and treatment of agreed uses as separately compensated subsidiary rights.[81]

> **Model clause: Reservation of AI rights**
>
> *Except to the extent expressly granted in this Agreement, the Author reserves all rights to use, authorise or license the Work for training, fine-tuning, evaluation, retrieval, grounding or other development or operation of a Generative AI System, and all rights to authorise use of the Work in AI-generated outputs.*
>
> *No general grant relating to digital, electronic, future or subsequently developed media shall be construed as granting an AI-related right unless the relevant use is expressly identified together with its scope, term, territory and consideration.*

6.4 Representations, Warranties and Delivery Certification

The author's warranties should address identifiable conduct within the author's control. They should avoid making the author the insurer of undisclosed conduct by a model developer. The publisher may require a higher standard where the author selects an unapproved tool, disregards warning signs or submits content without review.

> **Model clause: AI-related representations and warranties**
>
> *The Author represents and warrants that: (a) the Author has made the disclosures required by this Agreement; (b) to the Author's knowledge after reasonable review and verification, the Work does not contain AI-Generated Material that substantially reproduces identifiable third-party content in a manner that infringes any third-party right; (c) the Author has not knowingly or recklessly incorporated fabricated citations, quotations or attributions; (d) the Author has not uploaded confidential information, personal data or third-party content to a Generative AI System without lawful authority; and (e) the Author has complied with the applicable terms of use of each Generative AI System used in preparing the Work.*
>
> *The warranties in this clause shall be assessed by reference to the information and tools reasonably available to the Author at the date of delivery. The Author does not warrant the legality of a Generative AI System's undisclosed training practices unless the Author knew of, authorised or participated in the relevant conduct.*

A delivery certificate can improve record-keeping. It may attach an AI-use schedule listing the system, date, purpose, affected sections and review undertaken. The schedule should be treated as confidential editorial material. It may later assist the parties in responding to a complaint, demonstrating an internal review process and identifying the model version involved.

81 https://authorsguild.org/advocacy/artificial-intelligence/ai-model-clauses/

Model clause: Delivery certificate

On delivery of the final manuscript, the Author shall provide a signed AI Use Statement confirming whether Material AI Use occurred and, where applicable, completing the schedule attached to this Agreement. The Publisher shall use that statement for rights management, editorial review and claims handling and shall not publish it without the Author's consent unless required by law or reasonably necessary to address a third-party claim.

6.5 Indemnities, Liability Caps and Insurance

An AI-related indemnity should be tied to breach, causation and a defined category of loss. The mere presence of AI-assisted material should not trigger an indemnity. Under Indian law, sections 124 and 125 of the Indian Contract Act, 1872 address contracts of indemnity and the rights of an indemnity-holder when sued, while sections 73 and 74 govern compensation for breach and stipulated sums.[82] The drafting should state when a claim arises, who controls the defence and how settlements are approved.

Model clause: Author indemnity

The Author shall indemnify the Publisher against direct losses, damages, reasonable external legal costs and amounts paid under a settlement approved in accordance with this clause, to the extent arising from a third-party claim caused by the Author's material breach of the AI-related representations, warranties or restrictions in this Agreement.

The indemnity shall not apply to the extent that the claim results from: (a) a modification made by the Publisher without the Author's approval; (b) use of the Work outside the rights granted; (c) use of a Generative AI System selected or mandated by the Publisher, except to the extent the Author misused that system; or (d) the Publisher's failure to take reasonable steps to mitigate the loss after receiving notice.

The Publisher shall give prompt written notice of the claim, provide reasonable cooperation at the Author's cost, and permit the Author to participate in the defence. The Publisher shall not settle a claim in a manner that admits wrongdoing by, or imposes a non-monetary obligation on, the Author without the Author's prior written consent, such consent not to be unreasonably withheld.

The liability cap should reflect expected exposure, bargaining power, insurance and the price paid under the agreement. Publishing contracts often exclude intellectual property and confidentiality claims from a general cap. A complete exclusion may expose an individual author to liability far beyond the economic value of the arrangement. A negotiated super-cap, exclusions for deliberate misconduct, and access to media liability or professional indemnity coverage may produce a more workable allocation.

82 Indian Contract Act, 1872, ss 73, 74, 124 and 125.

Source of risk	Suggested allocation
Publisher mandates a particular AI tool	Publisher bears procurement and tool-selection risk. The author remains responsible for misuse and verification within the agreed workflow.
Author independently uses an unapproved tool	Author bears consequences of the unauthorised use, subject to causation and the agreed cap.
Tool is approved jointly	Each party bears matters within its control. Third-party contractual protection should flow to the party facing the claim.
Publisher alters material after delivery	Author should have no liability for the publisher's independent alteration.
AI provider offers an IP indemnity	The customer should review scope, exclusions, claim control, output filters and whether authors or affiliates receive any benefit.

6.6 Licensing Content for AI Training and Retrieval

A licence should describe the technical and commercial use with more precision than the expression "AI training". Depending on the architecture, a developer may seek rights for pre-training, fine-tuning, embeddings, retrieval-augmented generation, evaluation, safety testing, synthetic data creation, output display or a combination of these activities. Each activity may engage a different right, risk and revenue model.

Under Indian copyright law, reproduction includes storage in an electronic medium. Whether a particular training process infringes, falls within an exception or requires a licence remains fact-sensitive and unsettled. Section 52 contains enumerated exceptions, including fair dealing for specified purposes, and does not presently contain a general commercial text and data mining exception expressed in the terms used by certain foreign statutes.[83] A licensor should therefore avoid language that assumes all technical uses are legally equivalent.

However, this received significant judicial consideration in *ANI Media Private Limited v. OpenAI OpCo LLC*. At the interim stage, the Delhi High Court held that OpenAI's storage of ANI's literary works for training the large language models underlying ChatGPT fell, on a prima facie assessment, within "private or personal use, including research" under Section 52(1)(a) of the Copyright Act, 1957. The Court treated model training as an intermediate and closed process in which stored literary material is screened, organised and converted into machine-readable inputs for the iterative development of the model. Applying the doctrine of updating construction, the Court considered research capable of extending to machine learning undertaken at the instance of, and for the benefit of, human beings.[84]

The Court also rejected the proposition that a commercial purpose, by itself, excludes reliance on Section 52(1)(a). It observed that the expression "private" is capable of applying to a closed group or private company and is not confined to an individual. On the facts before it, the training material was stored in a closed environment, was accessible to the models, and was not made available to the public for access or download. The Court accordingly treated the use as private, although undertaken by a commercial enterprise.[85]

---

83 Copyright Act, 1957, ss 14 and 52. See also Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright, Part I: One Nation, One License, One Payment (December 2025), pp 13 to 22, https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf (accessed 27 July 2026).

84 ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024, paras 211 to 219.

85 Id., paras 179 to 190 and 208 to 211.

The purpose inquiry did not conclude the analysis. Section 52(1)(a) continued to require fair dealing. For the circumstances before it, the Court assessed fairness by considering: first, whether the copyrighted material was used only for model training; second, whether the use created economic competition with the copyright owner or caused actual or potential prejudice to its legitimate commercial interests; and third, whether the functions performed by the resulting system served the public interest. The Court found no demonstrated memorisation or regurgitation of ANI's works, no evidence that ChatGPT operated as a market substitute for ANI's news service, and no material showing a loss of subscribers or market share. The public benefits attributed to the technology, including research, education, translation, accessibility and dissemination of information, were also treated as relevant to the fairness assessment.[86]

The decision does not establish that every use of copyrighted material for AI training is fair dealing. The findings were expressly prima facie and confined to the application for interim relief. The Court emphasised the evidence before it, including the absence of demonstrated memorisation, substantially similar output, market substitution or use of ANI's material for a purpose apart from training. The order therefore provides a fact-sensitive framework rather than a general exemption for AI developers.[87]

6.6.1 Chain of title and author participation

The publisher must confirm that it controls the rights it proposes to license. Publication rights, electronic rights, database rights and AI-training rights may belong to different parties or may have been granted for limited purposes. Older contracts may not address machine training. The licence should identify the chain of title, reservations, territorial limitations, author approval requirements, collective management arrangements and any obligation to share proceeds.

6.6.2 Model training licence suite

> **Model clause: Grant and purpose**
>
> *Subject to the terms of this Agreement, the Licensor grants the Licensee a [non-exclusive] [non-transferable] licence during the Term and in the Territory to reproduce the works identified in Schedule 1 solely for the Permitted Purpose.*
>
> *"Permitted Purpose" means [pre-training / fine-tuning / retrieval-augmented generation / benchmarking / safety evaluation] of the Model identified in Schedule 2. No right is granted for any other model, affiliate, customer or downstream product unless expressly stated in Schedule 2.*
>
> *The licence does not authorise public display, communication, distribution, adaptation, translation, synthetic data creation or use in model outputs except as expressly set out in this Agreement*.[88]

[86] *Id., paras 234 to 256.*

[87] *Id., paras 271 to 274.*

[88] https://authorsguild.org/advocacy/artificial-intelligence/ai-model-clauses/

Model clause: Dataset integrity and access controls

The Licensee shall maintain a verifiable inventory of the Licensed Corpus, the date and method of ingestion, the Model version for which each item was used, and all persons and subcontractors given access.

The Licensee shall implement reasonable technical and organisational measures to prevent unauthorised access, extraction or onward use. Subcontractors may receive access only under written obligations no less protective than this Agreement, and the Licensee shall remain responsible for their acts and omissions within the scope of the licence.[89]

Model clause: Output safeguards

The Licensee shall maintain and periodically test safeguards designed to reduce the risk that the Model reproduces a substantial or commercially substitutive portion of a Licensed Work in response to an end-user prompt.

Where the Model provides a quotation, extract or retrieval result authorised by this Agreement, the output shall include the attribution, source link and usage limits specified in Schedule 3.

Upon receiving a substantiated complaint concerning a Licensed Work, the Licensee shall preserve relevant records, investigate the reported output and apply proportionate mitigation, which may include output suppression, retrieval filtering or temporary suspension pending review.[90]

The importance of technical record-keeping is demonstrated by *ANI Media Private Limited v. OpenAI OpCo LLC*, where the Court considered the training cut-off dates of the relevant models, the publication dates of the claimant's articles, the distinction between model training and retrieval-augmented generation, and the absence of demonstrated memorisation or regurgitation. A licence should therefore require records capable of identifying when a work was ingested, the model version for which it was used, whether the work is accessible through retrieval, and the safeguards applied against substantially similar output.[91]

Model clause: Audit and reporting

Within [30] days after each calendar quarter, the Licensee shall provide a report containing the information specified in Schedule 4, including ingested works, active Model versions, retrieval or display events where measured, complaints and material security incidents.

Once in each contract year, and additionally following reasonable evidence of material non-compliance, the Licensor may appoint an independent auditor bound by confidentiality to verify compliance. The audit shall be conducted in a manner that protects the Licensee's trade secrets, system security and information relating

[89] https://authorsguild.org/advocacy/artificial-intelligence/ai-model-clauses/

[90]

[91] *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr., Delhi High Court, interim order dated 24 July 2026, CS(COMM) 1028/2024, paras 82 to 84, 241 to 243 and 271 to 274.*

to other licensors. The Licensee shall bear the reasonable audit cost if the audit identifies a material underpayment or material breach.[92]

6.6.3 Termination and model survival

Termination presents a technical and contractual problem. Deleting source files is different from reversing their influence on model weights. The agreement should define the remedy that can realistically be performed and verified. An unconditional commitment to remove every learned influence may be incapable of objective proof.

Model clause: Effect of expiry or termination

On expiry or termination, the Licensee shall cease new ingestion and fine-tuning using the Licensed Corpus and shall delete or return source copies, embeddings and derivative datasets except to the extent retention is required by law or for the limited purpose of defending a claim.

The treatment of Model weights created before expiry shall be as selected in Schedule 5: (a) continued use subject to ongoing output restrictions and payment; (b) cessation of distribution of the affected Model version; or (c) commercially reasonable machine-unlearning or equivalent mitigation, where technically feasible and independently verifiable.

No provision shall require a technical outcome that the parties have not defined by measurable acceptance criteria. Confidentiality, accrued payment, audit, claims handling and dispute-resolution provisions shall survive.[93]

6.7 Royalty and Revenue-Sharing Models

Public announcements since 2023 confirm a developing market for licensed publisher content. OpenAI's December 2023 announcement regarding Axel Springer described the use of selected content in responses with attribution and links, use of publisher content for model training, and support for the publisher's AI ventures.[94] Many commercial terms remain confidential. The reported examples should therefore be used as evidence of transaction structures and market direction, rather than as a definitive pricing benchmark.

It is noteworthy that the availability of a potential fair-dealing defence does not remove the commercial role of licensing. In *ANI Media Private Limited v. OpenAI OpCo LLC*, the Delhi High Court's findings were made at the interim stage and depended on the manner and purpose of use demonstrated on the record. A contractual licence may provide rights and operational benefits extending beyond the statutory defence, including structured access, continuous data feeds, use of access-controlled material, retrieval and display rights, attribution standards, output permissions, usage reporting and contractual certainty. The commercial value of an AI-content licence should therefore be assessed by reference to the rights, data quality and risk allocation it supplies, rather than on the assumption that every act of model training necessarily requires permission.

The interim proceedings also illustrate that AI-content licences may have an ascertainable market value. In declining to grant an interim injunction, the Delhi High Court noted OpenAI's

[92] https://authorsguild.org/advocacy/artificial-intelligence/ai-model-clauses/
[93] https://authorsguild.org/advocacy/artificial-intelligence/ai-model-clauses/
94 https://openai.com/index/axel-springer-partnership/

submission that ANI had itself offered to license its digital news content for a fee of USD 7.5 million in a communication dated 3 October 2024. The Court observed that this indicated ANI's claim was capable of monetary valuation and that, should ANI ultimately succeed, compensation in damages would constitute an adequate remedy, whereas an injunction could significantly disrupt OpenAI's operations. Although these observations were made only for the purpose of deciding interim relief and do not determine the merits of the copyright dispute, they nevertheless demonstrate that negotiated licence values may serve as evidence of the commercial worth of AI-training rights in appropriate cases.

A royalty structure should match the technical use that can be measured. Per-output accounting may be workable where the system retrieves or displays an identified source. It is considerably harder where content contributes diffusely to model weights and no reliable attribution method connects an output to a particular work. Contracts should avoid payment formulas that depend on data the licensee cannot generate consistently.

Model	Use case	Principal drafting issue
Fixed corpus fee	One-time use of a defined archive for a defined model.	Scope creep, model updates and downstream use.
Annual licence fee	Continuing access to current content and periodic updates.	Renewal, inflation adjustment and change in product scope.
Minimum guarantee plus variable fee	Commercial product with uncertain adoption.	Definition of revenue, audit rights and recoupment.
Per-work ingestion fee	Clearly identified titles or datasets.	Does not capture downstream model value.
Retrieval or display fee	Output shows, quotes or links to identifiable content.	Reliable event logging, attribution and bot traffic.
Revenue share	Licensed corpus materially supports a defined product.	Attribution of revenue, deductions, bundling and affiliate transactions.
Collective blanket licence	Large repertoire with numerous rightsholders.	Mandate, repertoire data, tariff-setting and distribution methodology.

The agreement should define gross revenue, permitted deductions, bundled products, affiliate transactions, currency conversion, taxes, reserves, reporting periods and audit thresholds. Where authors participate in receipts, the publisher should specify whether the payment is treated as subsidiary-rights income, a royalty, or another category under the author agreement.

6.8 Liability Across the AI Value Chain

Liability should not be allocated solely by describing a party as a developer, deployer, publisher or user. The relevant inquiry is which party performed, authorised or controlled the act in question, what information it possessed, what contractual obligation it assumed, and whether the alleged loss was caused by its conduct. Current litigation also distinguishes between acquisition and storage of training material, the training process, and allegedly infringing outputs. Those issues may produce different outcomes on the same technical system.[95]

95 https://www.nortonrosefulbright.com/en/knowledge/publications/ce8eaa5f/ai-in-litigation-series-an-update-on-ai-copyright-cases-in-2026; https://mofotech.mofo.com/topics/ai-trends-for-2026---copyright-litigation-shifts-from-training-data-to-ai-outputs

Actor	Matters commonly within its control	Contractual tools
Model developer	Dataset acquisition, training architecture, model documentation and model-level safeguards.	Provenance warranties, audit material, IP indemnity, model cards, incident duties.
Platform deployer	Access controls, interface design, filters, notices, complaint handling and deployment settings.	Acceptable-use terms, logging, takedown, moderation standards, service levels.
Publisher	Tool procurement, author disclosures, editorial review and publication decisions.	AI policy, author warranties, approval rights, editorial verification, insurance.
Author or user	Inputs, prompts, supplied materials, verification and final submission.	Disclosure, restricted inputs, review certification, usage records.
Licensing intermediary	Authority to license, repertoire records, usage reporting and royalty distribution.	Mandate warranties, repertoire schedule, audit, claims reserve, distribution rules.

The European Union has enacted AI-specific regulatory duties through Regulation (EU) 2024/1689. Article 53 requires providers of general-purpose AI models to maintain a policy to comply with Union copyright law, including rights reservations under the Copyright in the Digital Single Market framework, and to publish a sufficiently detailed summary of training content.[96] These duties influence compliance and evidence. They do not create a complete civil-liability code for every copyright dispute involving developers, deployers and users.

The contractual allocation should also recognise limits imposed by applicable law. A party may allocate economic responsibility through indemnity without eliminating liability owed directly to a claimant or regulator. The agreement should state that internal allocation does not prejudice third-party rights and should require cooperation where several actors are named in the same proceeding.

6.9 Collective Licensing and Copyright Societies

Collective licensing may reduce the transaction cost of negotiating rights across a large catalogue. In India, Chapter VII of the Copyright Act regulates copyright societies. Section 33 addresses the business of issuing or granting licences, subject to the statutory framework and the right of an owner to license its own works consistently with its obligations as a member.[97] Any extension into AI-training licensing would require a clear mandate, accurate repertoire data and governance suitable for a technically complex use.

The Indian Performing Right Society and the Indian Reprographic Rights Organisation operate in different repertoires and should not be treated as interchangeable. A society or rights organisation cannot safely license AI-related rights merely because it administers another form of exploitation. Its governing documents, member mandates, chain of title and statutory registration must cover the relevant rights and class of works.

The DPIIT Working Paper on Generative AI and Copyright, Part I, published in December 2025, examines direct licensing, text and data mining exceptions, collective licensing and

96 Regulation (EU) 2024/1689 (Artificial Intelligence Act), art 53, Official Journal of the European Union, 12 July 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed 27 July 2026).
97 Copyright Act, 1957, Chapter VII, particularly s 33, Copyright Office, Government of India, https://copyright.gov.in/Copyright_Act_1957/chapter_vii.html (accessed 27 July 2026).

extended collective licensing. Its proposed policy framework includes a mandatory blanket licence and a proposed Copyright Royalties Collective for AI Training, identified by the acronym CRCAT.[98] The paper invited stakeholder comments and remains a policy proposal rather than enacted law. This distinction matters: the proposal remains under active government consideration, and the position described here may evolve if the Government ultimately adopts a different model.

Institutional issue	Question for policy and contract design
Mandate	Have members expressly authorised training, retrieval and output uses?
Repertoire	Can the organisation identify each work, rightsholder, term and restriction?
Tariff	Who sets the rate, by what criteria, and with what review mechanism?
Usage evidence	What records must developers provide and how will confidentiality be protected?
Distribution	How will revenue be allocated where model-level contribution cannot be traced per work?
Claims and conflicts	How are ownership disputes, duplicate mandates and non-member works handled?

6.10 Governing Law, Dispute Resolution and Evidence

The dispute clause should reflect the remedies and evidence likely to be required. A disagreement over royalty reporting may be suitable for audit and expert determination. Unauthorised dataset use, disclosure of an unpublished manuscript or release of a model outside the licence may require urgent injunctive relief. A single generic arbitration sentence rarely addresses both categories effectively.

6.10.1 Escalation and expert determination

Model clause: Escalation and technical expert

A party shall notify the other of a dispute with reasonable particulars. Designated senior representatives shall meet in good faith within [10] Business Days and attempt resolution for [15] Business Days.

A dispute confined to usage measurement, royalty calculation, technical feasibility of deletion or machine unlearning, or compliance with agreed technical criteria may be referred to an independent expert jointly appointed by the parties. The expert shall act as an expert and not as an arbitrator. Questions of law, infringement, validity, fraud and contractual interpretation remain for the court or arbitral tribunal.

An expert determination shall not prevent either party from seeking urgent interim relief or preserving evidence.

98 Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright, Part I: One Nation, One License, One Payment (December 2025), chapters 4 and 5 and list of abbreviations, https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf (accessed 27 July 2026). The paper describes CRCAT as the Copyright Royalties Collective for AI Training.

6.10.2 Arbitration and interim relief

For an India-seated arbitration, the agreement should be drafted with the Arbitration and Conciliation Act, 1996 in view. Sections 9 and 17 provide routes for interim measures through courts and arbitral tribunals respectively.[99] The clause should identify the seat, institution or ad hoc procedure, rules, language and number of arbitrators. Cross-border arrangements may justify institutional arbitration, particularly where confidentiality, emergency relief and enforceability are material.

Model clause: Arbitration

This Agreement shall be governed by the laws of India. Any dispute arising out of or in connection with this Agreement, including its existence, validity, interpretation, performance, breach or termination, that is not resolved under the escalation procedure shall be finally resolved by arbitration administered by [institution] under its rules in force when the arbitration is commenced.

The seat and venue arbitration shall be [for ex. New Delhi / Mumbai / Singapore]. The tribunal shall consist of [one / three] arbitrator[s]. The language shall be English. The award shall be final and binding.

Nothing in this clause prevents a party from seeking urgent interim, conservatory or injunctive relief from a court of competent jurisdiction or, where available, an emergency arbitrator. The parties shall maintain the confidentiality of the proceedings, evidence and award, subject to disclosure required by law, enforcement or professional advice.

6.10.3 Evidence preservation

AI disputes are particularly dependent on contemporaneous technical records. The agreement should establish a proportionate retention regime before a complaint arises. The objective is traceability, while respecting privacy, trade secrets and data-minimisation duties.

Model clause: Preservation and cooperation

Each party shall preserve records reasonably necessary to verify compliance with this Agreement, including applicable dataset inventories, ingestion records, model and version identifiers, author AI-use statements, prompts and outputs relevant to a notified complaint, retrieval logs, licence records, usage statements and royalty reports.

Upon receiving notice of a claim, audit or reasonably anticipated dispute, each party shall suspend routine deletion of relevant records and preserve them subject to appropriate confidentiality, privilege, security and personal-data safeguards.

The parties shall cooperate in good faith to identify the source of a disputed output, implement proportionate mitigation and respond to lawful requests. Cooperation does not require disclosure of legally privileged material or unrelated trade secrets.

99 Arbitration and Conciliation Act, 1996, ss 9 and 17.

6.11 Drafting Checklist

Before an AI-related publishing or content licence is signed, counsel should confirm the following matters:

- Identify the AI system, model, service and version to which the agreement applies.
- Distinguish routine assistance, material use and restricted use.
- Define the author's disclosure, verification and delivery obligations.
- Regulate the publisher's use of manuscripts and author information in third-party systems.
- State whether AI-related rights are reserved, granted or treated as subsidiary rights.
- Confirm chain of title and any author, agent, society or third-party approval.
- Separate training, fine-tuning, retrieval, evaluation, synthetic data and output-display rights.
- Define output safeguards, attribution, source links and complaint handling.
- Allocate tool-selection risk to the party controlling procurement and configuration.
- Tie warranties and indemnities to information, control, breach and causation.
- Address liability caps, deliberate misconduct, confidentiality, intellectual property and insurance.
- Select a payment model that can be measured and audited.
- Define the treatment of source files, embeddings and model weights after termination.
- Require technical and financial records adequate for audit and dispute resolution.
- Provide escalation, expert determination, arbitration or court jurisdiction suited to the dispute.
- Preserve urgent relief for unauthorised use, confidentiality breaches and imminent model release.
- Include a change-in-law mechanism and require periodic review of the AI schedule.

6.12 Conclusion

AI contracting in publishing should proceed from control, evidence and commercial purpose. The author controls the process used to prepare and verify the manuscript. The publisher controls procurement, editorial acceptance and publication. The developer controls dataset engineering and model-level safeguards. A rights organisation controls the accuracy of its mandate and repertoire. The agreement should reflect those areas of control and should create records capable of demonstrating compliance.

A durable contract will define the relevant technology and permitted use, reserve rights that have not been priced, allocate risk by reference to breach and causation, and establish remedies that remain technically achievable. It should also anticipate the practical course of a dispute: preservation of model and dataset records, rapid mitigation of suspect outputs, access to urgent relief, and a forum capable of handling legal and technical evidence. These features give parties a workable framework while legislation, regulatory policy and case law continue to develop.

Chapter VII:
Policy Recommendations and the Road Ahead



The preceding chapters have traced a single inquiry across several points of contact between artificial intelligence and the publishing sector: training, output, integrity, and contract. India's courts and regulators have engaged with each of these questions in turn, through interim rulings, active consultation, and initiatives such as the IndiaAI Mission, though a unified statutory framework has yet to emerge. That is unsurprising for a technology moving this quickly; most major jurisdictions face the same difficulty. This chapter sets out the authors' assessment of where matters presently stand, the priorities for policy development, and, more immediately, the steps publishers can take under the law as it exists today.

For the publishing industry, the calculus is not simply one of risk. Generative AI offers genuine commercial and editorial opportunity, in workflow efficiency, discoverability, and new forms of licensing revenue, and the recommendations below are not intended to discourage that opportunity. They are intended to help publishers pursue it on terms that are commercially sound and legally defensible.

7.1 Where the Law Stands

Three propositions follow from the discussion in the preceding chapters. First, the legality of training AI models on copyrighted works under Indian law remains open. Such judicial guidance as exists is confined to interim observations in ongoing proceedings; no court has yet ruled conclusively on whether, or on what conditions, Section 52 of the Copyright Act, 1957[100] accommodates such use. Second, the copyright status of AI-assisted output is similarly unsettled. Section 2(d)(vi)[101] predates large language models by decades and has not been tested against them; its application to a work substantially generated, rather than merely assisted, by AI remains a live question of statutory interpretation. Third, the plagiarism-detection infrastructure on which publishers and universities presently rely was built to catch textual overlap, and is correspondingly weaker at identifying undisclosed AI assistance or competent paraphrase – a limitation examined in Chapter 4.

None of this amounts to a legislative vacuum; it is, more accurately, an area of active judicial and regulatory development that has not yet settled. Pending that settlement, the contractual protections discussed in Chapter 6 remain the most reliable tool available to publishers.

7.2 Legislative and Policy Priorities

The absence of settled law should not be read as an absence of engagement. The IndiaAI Mission[102], the DPIIT's ongoing consultation process[103], and the 2026 amendments to the IT Rules[104] together indicate a policy apparatus that is actively working through these questions rather than one that has been slow to act. What follows are areas where, in the authors' view, further clarity would benefit the sector – not criticisms of the pace of that work.

---

[100] *The Copyright Act, 1957, s. 52(1)(a).*
[101] *Copyright Act, 1957, ss 2(d)(vi), 14, 17 to 19 and 30, Copyright Office, Government of India, https://copyright.gov.in/documents/copyright_act_1957.pdf (accessed 27 July 2026).*
[102] *https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2012375&reg=48&lang=2; https://www.pib.gov.in/PressReleasePage.aspx?PRID=2178092&reg=3&lang=2*
[103] *Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright – Part I: One Nation One License One Payment, December 2025: https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf (accessed 27 July 2026).*
[104] *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified 10 February 2026 (G.S.R. 120(E)), amending the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, and brought into force on 20 February 2026.*

Two provisions merit particular legislative or judicial attention: the scope of Section 52 as applied to AI training, and the interpretation of Section 2(d)(vi) as applied to AI-assisted authorship. Clarity on either, whether by amendment or judicial pronouncement, would materially reduce transactional uncertainty for authors, publishers, and technology developers alike.

The DPIIT's Working Paper[105], proposing a blanket licensing framework for AI training administered through a Copyright Royalties Collective, is a policy proposal under consultation, not a settled government position, and is best treated as one considered contribution among several. Questions concerning an opt-out mechanism, the treatment of personality rights, and the position of open-source and share-alike licences remain genuinely contested, and reasonable stakeholders differ on each. Separately, the 2026 IT Rules amendments[106] have strengthened platform accountability; further clarification of the position of platforms that train AI models, as distinct from those that merely host content, would be a useful next step.

FICCI's submission during the consultation, discussed at Section 5.2, illustrates this point concretely: it welcomed transparency measures such as AI-developer registration while cautioning against replacing voluntary licensing with a mandatory blanket regime, reflecting the same tension between certainty of compensation and negotiated choice that runs through this debate.

None of these questions is capable of resolution by any single stakeholder. Government, industry bodies such as FICCI, publishers, technology developers, academic institutions, and creators each hold information the others lack, and a workable framework is more likely to emerge from structured dialogue among them – of the kind already under way through DPIIT's consultation process – than from any single interest being imposed on the rest.

7.3 What Publishers Can Do Now

The absence of finalised legislative guidance is not a reason to defer action. The measures below can be adopted under the law as it presently stands, and are intended to remain fit for purpose regardless of how the legislative position develops.

At the contractual level, publishers should consider incorporating AI-specific provisions into standard author and publishing agreements, rather than addressing these questions only once a dispute has arisen. Such clauses should: define "Material AI Use" as distinct from routine editorial assistance; require disclosure of any Material AI Use by the author; require the author to confirm that AI-generated citations, quotations, and factual assertions have been independently verified; and address rights relating to AI training, fine-tuning, and retrieval expressly, rather than by reliance on broad or historic grants of digital or future-media rights. Warranties and indemnities should be limited to matters within the author's own knowledge and control, and should not extend to the undisclosed training practices of third-party AI developers.

At the editorial stage, publishers may require an AI Use Statement with each manuscript, recording the system used, its purpose, the sections affected, and the review undertaken. Editors

[105] *Department for Promotion of Industry and Internal Trade, Working Paper on Generative AI and Copyright – Part I: One Nation One License One Payment, December 2025: https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf (accessed 27 July 2026).*

[106] *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified 10 February 2026 (G.S.R. 120(E)), amending the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, and brought into force on 20 February 2026.*

and peer reviewers should treat AI-detection and similarity scores as an indicator warranting further inquiry, not as conclusive evidence of misconduct, given the limitations discussed in Chapter 4.

In licensing, agreements for AI training or retrieval use should establish a clear chain of title, define permitted use with precision, and include audit and reporting rights sufficient to identify the works used and the safeguards applied. Provision should also be made for the treatment of source files, embeddings, and model weights on termination, since deletion of source material does not necessarily remove its influence from an already-trained model.

For collecting societies, should collective licensing frameworks for AI training be adopted, collecting societies will need to strengthen repertoire records and governance to administer them credibly, and will require a clear and specific mandate from rights holders before licensing AI-related uses. Publishers and collecting societies that establish sound contractual and administrative practice now will be better placed to respond as the legislative position develops.

Collaboration also has a place among these immediate steps, not only in longer-term policy design. Publishers, individually and through industry bodies such as FICCI, can engage constructively with DPIIT's ongoing consultation process, share practical experience on licensing and disclosure with technology developers and academic institutions, and coordinate with creators' associations on issues of shared concern. Stakeholders who take part in this dialogue now are better placed to help shape a workable framework, rather than respond to one designed without their input.

7.4 The Road Ahead

The framework governing AI and copyright will continue to develop, in India and elsewhere, through the combined work of courts, regulators, and industry rather than through any single decisive event. The DPIIT Working Paper remains under consultation; the pending appeal in *Thomson Reuters v Ross Intelligence*[107] and the implementation of Article 53 of the EU AI Act[108] may in time offer persuasive reference points for Indian courts and policymakers, alongside domestic developments through the IndiaAI Mission. Litigation on AI training and copyright remains pending in India and elsewhere, and the applicable principles are still being worked out through that process.

In this environment, practical governance matters more, not less. Publishers who maintain clear records, adopt carefully drafted AI-specific provisions, and require transparent disclosure of AI use will be better placed to manage legal and commercial risk while the framework develops, and will generate the kind of practical experience that informs sound policy. These measures cannot resolve the underlying legal questions; they can, however, provide certainty and accountability in the interim.

Approached in this spirit, the challenges generative AI presents for copyright and publishing are also an invitation: to reaffirm why human creativity, judgment, and accountability remain indispensable, and to build a publishing ecosystem in which technology strengthens, rather than displaces, the value that authors, editors, and publishers bring to the written word.

[107] *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 25-2153 (3d Cir.) (appeal pending).*
[108] *Regulation (EU) 2024/1689 (Artificial Intelligence Act), art 53, Official Journal of the European Union, 12 July 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed 27 July 2026).*

Navigated with care and in good faith by all concerned, this moment holds real promise for a more dynamic, inclusive, and rewarding future for creativity and publishing alike.

The opportunities generative AI offers publishers – in efficiency, discoverability, and new licensing revenue – are best realised alongside sustained attention to their implications for creators' rights and research integrity, not instead of it. That balance depends on continued, good-faith collaboration between government, industry bodies such as FICCI, publishers, technology companies, academic institutions, and creators, each contributing expertise the others cannot supply.

About S.S. Rana & Co. S. S. Rana & Co.

Established on September 1, 1989, is a premier full-service law firm in India, renowned for its specialized expertise in Intellectual Property Law, Corporate & Commercial Laws, and Litigation, Arbitration & Dispute Resolution. The founding partners bring over 55 years of combined experience in IP litigation, having successfully represented several globally recognized brands in trademark infringement and passing-off matters before courts across India. The Firm has developed a strong global presence and serves as a single-window legal solutions provider for national and multinational corporations. S. S. Rana & Co. advises clients on IP filings, prosecution, enforcement, portfolio management, and complex legal issues across India and the SAARC region, including Bangladesh, Nepal, Pakistan, Bhutan, Sri Lanka, Maldives, Myanmar, Mongolia, and Afghanistan. Additionally, the Firm assists Indian and international clients with global IP filings and prosecution across multiple jurisdictions. Recognized as one of the best IP and corporate law firms in India, S. S. Rana & Co. has consistently been ranked and acknowledged by leading legal research and industry platforms such as The Legal 500, India Business Law Journal, World Trademark Review 1000, and Managing IP. Headquartered in New Delhi, the Firm operates through regional offices in Chennai , Noida, Mumbai, Bengaluru, and Kolkata, along with associate offices across the Indian subcontinent. With a diverse client base comprising Fortune 500 companies, multinational corporations, domestic enterprises, start-ups, and grassroots innovators, S. S. Rana & Co. delivers strategic, business-focused legal solutions to clients in India and worldwide.

ABOUT FICCI

Established in 1927, FICCI is the largest and oldest apex business organisation in India. Its history is closely interwoven with India's struggle for independence, its industrialization, and its emergence as one of the most rapidly growing global economies.

A non-government, not-for-profit organisation, FICCI is the voice of India's business and industry. From influencing policy to encouraging debate, engaging with policy makers and civil society, FICCI articulates the views and concerns of industry. It serves its members from the Indian private and public corporate sectors and multinational companies, drawing its strength from diverse regional chambers of commerce and industry across states, reaching out to over 2,50,000 companies.

FICCI provides a platform for networking and consensus building within and across sectors and is the first port of call for Indian industry, policy makers and the international business community.



Contact Us

S.S. Rana & Co.

81/2, Aurobindo Square, Aurobindo Marg, Adhchini, New Delhi 110017

T: 011 40123000

E: *info@ssrana.com*

Vikrant@ssrana.com

W: https://ssrana.in

FICCI

Federation House, Tansen Marg, New Delhi, 110001

T: +91-11-23487515

E: gaurav.gaur@ficci.com

aastha.gupta@ficci.com

W: https://ficci.in/