

# AI & Copyright

**Table of Contents**

I. Introduction .....	1
II. The Copyright Conundrum .....	1
III. What is Fair Use? .....	2
IV. Fair Use v. Fair Dealing .....	4
V. The Case of OpenAI and ChatGPT .....	4
VI. Conclusion .....	5

## I. Introduction

2023 was a big year for generative AI. AI - or artificial intelligence - suddenly went from being a thing of science fiction or a topic of debate among technophiles on online reddit forums, to being mainstream. When Microsoft backed OpenAI launched ChatGPT around the end of 2022, it took big tech by surprise. But unsurprisingly, it didn't take long for the likes of Google, Meta, and Amazon to jump on the AI bandwagon. As a result, we saw the rapid advancements in AI and adoption of AI tools by millions of daily users to assist them in all kinds of tasks. The technology has certainly wowed the world, even as discourse rages around whether the sudden advance of AI is to be celebrated or feared.

Perhaps the most widely used AI enabled tools are large language models or LLMs (such as ChatGPT and Bard), which have the ability to understand natural human language and generate human-like responses. These language models are trained on vast sets of data, usually billions of examples of written texts, and invariably, a majority of this data set comprises copyright protected written works.

In my previous article titled 'Regulating AI'<sup>1</sup> I had examined the potential risks associated with AI based LLMs and made recommendations for regulatory intervention. Among the risks highlighted, I had discussed the issue of copyright infringement and how current copyright laws do not address the challenge posed by LLMs.

Since then, a number of lawsuits have been filed which are likely to test the boundaries of copyright law. Most prominently, in December 2023, the New York Times sued<sup>2</sup> OpenAI and Microsoft for infringement of its copyright, contending that millions of articles published by the paper were used to train the automated chatbot. In similar actions, actress and comedian Sarah Silverman sued<sup>3</sup> OpenAI and Meta, and novelists and authors John Grisham, Jonathan Franzen and Elin Hilderbrand sued<sup>4</sup> OpenAI, both for copyright infringement, contending that their written copyrighted works were used by the companies, without permission, to train their AI chatbots. These are among dozens of other lawsuits which are now pending before US courts on the issue of copyright infringement by AI companies.

With these technological advances, is it perhaps time for a new scrutiny of existing copyright law? In the Indian context, does the Copyright Act, 1957 adequately protect the interests of authors and other creators of original literary works? This article broadly examines the existing copyright protection regimes, the defenses or exceptions to the use of copyright protected work, and how advances in technological innovations are posing new challenges for copyright law, keeping a lens on the dispute between OpenAI and the New York Times.

## II. The Copyright Conundrum

A significant portion of the data set used in training LLMs like ChatGPT and Bard comprises copyrighted written works. The collective content of all such copyrighted works is key to the responses generated by LLMs. Infact, OpenAI has publicly admitted<sup>5</sup> that it would be impossible to train its AI chatbot ChatGPT without access to copyright protected works.

LLMs do not comprehend the meaning conveyed by language. Instead, they generate responses based on learnt patterns and relationships between words and phrases. LLMs' responses do not simply copy-paste material from any individual written work. Rather, the sequence of words and

---

<sup>1</sup> <https://argus-p.com/papers-publications/thought-paper/regulating-ai/>

<sup>2</sup> <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>

<sup>3</sup> <https://www.nytimes.com/2023/07/10/arts/sarah-silverman-lawsuit-openai-meta.html>

<sup>4</sup> <https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html>

<sup>5</sup> <https://www.theguardian.com/technology/2024/jan/08/ai-tools-chatgpt-copyrighted-material-openai>

---

phrases identified in written works by multiple authors, are reorganized or re-sequenced by LLMs, to predict a response.

The typical defense offered by AI companies<sup>6</sup> freely using copyrighted works, i.e., without authorization or payment, is that such use of publicly available internet materials for training their AI based LLMs is “fair use” under copyright law.

On the flip side, authors and copyright holders argue that the responses generated by LLMs are a nuanced reproduction of the collective works of several copyright holders and hence, a form of infringement of intellectual property rights. Moreso, typically no credit is given to the authors they borrow from and LLMs offer a competing product that threatens the business of original copyright holders.

The question that thus arises is whether the use of copyright protected works in training of LLMs constitutes an infringement of copyright. Or does such use fall within the exception to the general rule, of a copyright holder’s exclusive rights, under the doctrine of fair use.

### III. What is Fair Use?

Fair use is the term used under US copyright law, akin to fair dealing under English or Indian copyright law. Broadly, fair use or fair dealing is the right to use a copyrighted work under certain conditions without permission of the copyright owner<sup>7</sup>. It is the thin line of difference between *bonafide*, legitimate use of a copyright protected work, as opposed to its reproduction or blatant copying<sup>8</sup>. Differently put, the doctrine of fair use or fair dealing acts as a limitation on the exclusive right of the holder of copyright as it permits the use of copyright protected work without the threat of a suit for infringement.

The fair use doctrine is perhaps the most significant limitation on copyright protection, developed out of judicial recognition that certain acts of copying are defensible when the public interest in permitting the copying outweighs the author’s interest in copyright protection<sup>9</sup>.

#### A. International Law

The first recognition of the doctrine of fair dealing in an international convention, as an exception to the rule of infringement of copyright, was under the Berne Convention for the Protection of Literary and Artistic Works, which was adopted in 1886 for recognition of copyright between signatory nations. The Berne Convention states that it shall be a matter for legislation in signatory or member countries to permit the reproduction of copyright protected works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author<sup>10</sup>.

In 1994, members of the World Trade Organization signed the TRIPS Agreement or the Agreement on Trade-Related Aspects of Intellectual Property Rights. The TRIPS Agreement is the most comprehensive multilateral agreement on intellectual property and it recognizes the doctrine of fair dealing in copyright protected works. The Agreement mirrors the words of the Berne Convention, stating that member countries shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict

---

<sup>6</sup> <https://openai.com/blog/openai-and-journalism>

<sup>7</sup> <https://ogc.harvard.edu/pages/copyright-and-fair-use>

<sup>8</sup> Sufiya Ahmed, *Fair Dealing in Indian Copyright Law*, Volume 26, Journal of Intellectual Property Rights, 96-102 (98), (2021).

<sup>9</sup> Benjamin Ely Marks, *Copyright protection, privacy rights, and the Fair Use Doctrine: The Post-Salinger decade reconsidered*, Volume 72, New York University Law Review, 1376-1419 (1377), 1997.

<sup>10</sup> Article 9(2) of the Berne Convention for the Protection of Literary and Artistic Works.

---

with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder<sup>11</sup>.

Most WTO member countries have since incorporated into their national laws the doctrine of fair use or fair dealing as an exception to the general rule of exclusive rights vested in the holder of a copyright.

## **B. US Law**

In the United States, the Copyright Act, 1976 recognizes the exclusive right of a copyright holder to reproduce, prepare derivative works based on, distribute, etc., the original literary work<sup>12</sup>. This exclusive right, however, is limited by the doctrine of fair use incorporated under Section 107 of the Act. In coming to a conclusion if a work qualifies as “fair use” of copyright protected written works, courts in the United States apply a four factor<sup>13</sup> test, viz. (i) the purpose and character of the use (i.e., for non-profit educational use, or for criticism, comment, news reporting, teaching, scholarship, or research); (ii) the nature of copyrighted works (i.e., published or unpublished, whether commercially available, fiction or nonfiction); (iii) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (iv) the effect of the use upon the potential market for or value of the copyrighted work.

## **C. English Law**

The concept of fair dealing in the UK was originally embodied in the UK Copyright Act, 1911, spelled out more fully in the UK Copyright Act, 1956 and later reinforced in the UK Copyright, Designs and Patents Act, 1988.<sup>14</sup> In the UK, fair dealing in copyright protected work is judged on whether (i) the usage falls within one of the specified categories of permitted use (research, private study, criticism or review, caricature, parody or pastiche does, illustration for instruction, educational use); (ii) the source has been adequately acknowledged; and (iii) the purpose is non-commercial and hence unlikely to impact the commercial interests of the original copyright holder.<sup>15</sup>

## **D. Indian Law**

Indian copyright law borrows heavily from English law. Section 13 of the (Indian) Copyright Act, 1957 protects original literary works by recognizing the author’s copyright and his exclusive right to do so himself or authorize, the reproduction, making of copies, translation or making adaptations, etc. of his original work. This exclusive right, though, is subject to Section 52 of the 1957 Act which incorporates the doctrine of fair dealing and lists out the standard exceptions or defenses to copyright infringement. The Act does not define fair dealing, but as per Section 52, a fair dealing in literary work for the purpose of (i) private or personal use, including research; (ii) criticism or review, whether of that work or of any other work; and (iii) the reporting of current events and current affairs, including the reporting of a lecture delivered in public; would not constitute an infringement of copyright. The proviso is Section 52 also mandates that title of the work must be identified along with due acknowledgement given to the holder of original copyright.

---

<sup>11</sup> Article 13 of the Agreement on Trade-Related Aspects of Intellectual Property Rights, signed on April 14, 1994 by WTO member countries.

<sup>12</sup> Section 106, (United States) Copyright Act, 1976.

<sup>13</sup> Section 107, (United States) Copyright Act, 1976.

<sup>14</sup> Lynette Owen, *Fair dealing: a concept in UK copyright law*, Volume 28, Journal of Scholarly Publishings, 229-231 (229), 2015.

<sup>15</sup> Sections 29, 30, 32 and 33 of the UK Copyright, Designs and Patents Act, 1988.

#### **IV. Fair Use v. Fair Dealing**

While largely similar and intended for the same purpose, i.e., to promote creativity and not allow copyright protections to stifle research and innovation, fair use in the US and fair dealing under English and Indian copyright law differ slightly. It is widely understood that fair use under American jurisprudence is a broader concept since the purposes listed out under Section 107 of the (US) Copyright Act, 1976 (criticism, comment, news reporting, teaching, etc.) are illustrative and not exhaustive. Thus, potentially, fair use under US copyright law can apply to any other purpose provided that the mandatory factors of Section 107, delineated above, are satisfied.

In contrast, the concept of fair dealing under English or Indian copyright law is more rigid or restrictive. The purposes for which a copyright protected work could be used, and such use could qualify as fair dealing is statutorily limited. In other words, for an action to be protected from the threat of infringement of copyright, it must strictly fall within the purpose limitation prescribed in the relevant English and Indian statutes. The purposes spelt out, under Indian and English copyright law, are also very similar and include research, criticism or review, reporting of current events, etc.

#### **V. The Case of OpenAI and ChatGPT**

OpenAI argues that training of its AI chatbot ChatGPT using publicly available internet materials is fair use. This includes the use of millions of articles published by the New York Times to train its AI model.

To test this argument on the parameters of US copyright law, one would have to first examine if OpenAI's use of this online material qualifies on the purpose limitation. Since the list of purposes spelt out in Section 107 is only illustrative and not exhaustive, it is foreseeable that OpenAI's purpose may qualify on the first criterion for fair use. Having said that, US courts have typically found in favor of non-profit purposes as qualifying for fair use and held against use which is of a commercial nature. Since ChatGPT is already being monetized by OpenAI, this is the first potential hurdle in OpenAI's defense of fair use.

Second, is the question of the copyrighted work itself, i.e., all articles of the New York Times, a result of decades of painstaking journalistic work. While these articles are published works and available commercially, the material is available subject to payment of monthly subscription fee, a factor that is likely to go against OpenAI's fair use argument. Add to this, US courts typically tend to protect creative works, and it is arguable that articles published in the paper are perhaps original and creative, in that they are not easy to replicate and a result of countless hours of labor, and simply given the reputation that the publisher in this case enjoys.

Third, is the question of how substantial is the use of the copyright protected work. This factor may not strictly apply in the case of OpenAI since the chatbot, in the normal course, although the New York Times alleges otherwise in its complaint, does not verbatim repeat the articles used in its training but transforms the text based on user query. However, given that the AI chatbot is trained on *all* articles ever published by the New York Times, this test is also likely to go against OpenAI's argument of fair use.

Last, is the question of the effect on the potential market value of the copyright protected work. It is easy to see how ChatGPT offers a competing product to that offered by the New York Times. ChatGPT is already being monetized, while the commercial value of the original works, forming the underlying data set, is, at least in theory, diminishing over time as more and more people turn to the simplified, summarized presentations of these works. Though it will be difficult, if not impossible, to compute specific damages stemming directly from the alleged infringement.

---

## ***Under (Indian) Copyright Act***

It is only a matter of time before OpenAI's defense of fair use, fair dealing in the Indian context and hence more restrictive, is tested under Indian law. As discussed above, Section 52 of the (Indian) Copyright Act prescribes a strict purpose limitation. Any use of copyrighted works for purposes other than (i) private or personal use; (ii) criticism or review; and (iii) reporting of current events and affairs, is deemed not to be fair dealing with such protected work. Given the novel purpose for which OpenAI is making use of copyrighted materials, which, at the face of it, is beyond the purposes spelt out in Section 52, OpenAI's argument does not pass the muster of Section 52 and the argument of fair dealing must fail.

Add to this, Open AI does not give acknowledgement to the authors it borrows from, which, under Indian law, is a necessary ingredient for an unauthorized use of copyrighted work to qualify as fair dealing. On this account as well OpenAI's argument of fair dealing must fail.

Another factor that may come into play is the concept or element of fairness, common to both fair use and fair dealing. Mere dealing with the work for the relevant purpose is not enough; it must also be dealing which is fair for that purpose whose fairness must be judged in relation to that purpose<sup>16</sup>. While entertaining claims of fair dealing, Indian courts have often employed the test of fairness. It is widely accepted in Indian, as well as English and American, jurisprudence that while key tenets of fair dealing cannot be ignored, there exists no universal rule or straightjacket formula. Every case of fair dealing has to be adjudicated on its own facts and what may be unfair in one context may be perfectly fair in another.

Therefore, when tested, the law may ultimately protect a copyright holder's rights, but given the evolving nature of AI technologies, and rampant unethical, to say the least, use of copyright protected works, the legal framework may have to adapt by carrying out necessary amendments to address emerging issues and balance competing rights. Perhaps time has come where amendment to copyright law is necessitated to give statutory recognition to new forms of infringement, one where machine learning may produce texts which are arguably "original" but are a mere nuanced reproduction/ amalgamation of information from multiple copyright protected works.

Fundamentally though, the dispute between the New York Times and OpenAI is about appropriately compensating content creators. One possible way to put the current controversy to rest is to mandate that companies like OpenAI can only train their AI chatbots on copyright protected works if they have the license or authorization from the copyright holders. In fact, media organizations are already striking licensing deals<sup>17</sup> with AI based tech companies that may prove to be mutually beneficial arrangements, and especially come to the aid of traditional forms of journalism like the print media, which is constantly seeing a decline in readership.

## **VI. Conclusion**

Training of LLMs is a novel concept that poses new challenges to copyright law. The novel arguments on both sides seek to expand copyright law into new territory, something that the law, as it was originally written, was not designed for.

So do OpenAI's actions infringe New York Times's copyright? Will OpenAI's use of published articles of the New York Times to train its AI chatbot ChatGPT qualify as fair use? While these questions may soon be answered in US court decisions, and sooner or later come before courts in India as well, one thing is certain that legal systems lag behind technological advancements in the AI space and the need for regulation is growing.

---

<sup>16</sup> Sufiya Ahmed, *Fair Dealing in Indian Copyright Law*, Volume 26, Journal of Intellectual Property Rights, 96-102 (98), (2021).

<sup>17</sup> <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>

**Contributed by:**



**Udit Mendiratta  
Partner**

## **DISCLAIMER**

This document is merely intended as an update and is merely for informational purposes. This document should not be construed as a legal opinion. No person should rely on the contents of this document without first obtaining advice from a qualified professional person. This document is contributed on the understanding that the Firm, its employees and consultants are not responsible for the results of any actions taken on the basis of information in this document, or for any error in or omission from this document. Further, the Firm, its employees and consultants, expressly disclaim all and any liability and responsibility to any person who reads this document in respect of anything, and of the consequences of anything, done or omitted to be done by such person in reliance, whether wholly or partially, upon the whole or any part of the content of this document. Without limiting the generality of the above, no author, consultant or the Firm shall have any responsibility for any act or omission of any other author, consultant or the Firm. This document does not and is not intended to constitute solicitation, invitation, advertisement or inducement of any sort whatsoever from us or any of our members to solicit any work, in any manner, whether directly or indirectly.

**You can send us your comments at:**

**[knowledgecentre@argus-p.com](mailto:knowledgecentre@argus-p.com)**

Mumbai | Delhi | Bengaluru | Kolkata

[www.argus-p.com](http://www.argus-p.com)



## Recent Papers/ Articles



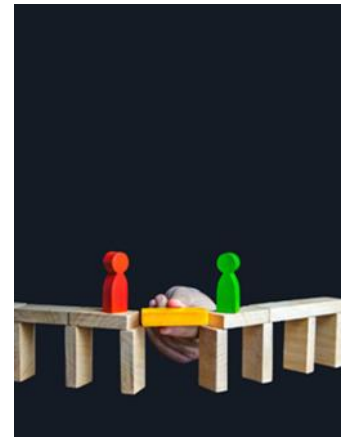
November 2023

**Green Hydrogen (Series 2) - A Practical Insight Into Green Hydrogen Project Development**  
Energy and Infrastructure



November 2023

**Angel Tax And The (Un)Ease Of Doing Business In India**  
Corporate and M&A,  
Private Equity & Venture  
Capital



August 2023

**The Exclusive Jurisdiction of Courts regarding Wills - Scope for arbitration/ mediation of disputes arising out of a Will**  
Disputes



July 2023

**Green Hydrogen (Series 1) : Framework And Developments In India**  
Energy and Infrastructure



May 2023

**Regulating AI**  
Technology and Data  
Privacy



May 2023

**Navigating Crossroads Of IBC And RERA**  
Corporate Restructuring  
& Insolvency

For more Papers/ Articles [click here.](#)

#### **MUMBAI**

11, Free Press House  
215, Nariman Point  
Mumbai 400021  
T: +91 22 6736 2222

#### **DELHI**

7A, 7th Floor, Tower C, Max  
House, Okhla Industrial Area,  
Phase 3, New Delhi 110020  
T: +91 11 2370 1284/5/7  
T: +91 11 45062522

#### **KOLKATA**

Binoy Bhavan  
3rd Floor, 27B Camac Street  
Kolkata 700016  
T: +91 33 40650155/56

#### **BENGALURU**

68 Nandidurga Road  
Jayamahal Extension  
Bengaluru 560046  
T: +91 80 46462300