

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS



Open Access, Refereed Journal Multi-Disciplinary
Peer Reviewed

www.ijlra.com

DISCLAIMER

No part of this publication may be reproduced, stored, transmitted, or distributed in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Managing Editor of the *International Journal for Legal Research & Analysis (IJLRA)*.

The views, opinions, interpretations, and conclusions expressed in the articles published in this journal are solely those of the respective authors. They do not necessarily reflect the views of the Editorial Board, Editors, Reviewers, Advisors, or the Publisher of IJLRA.

Although every reasonable effort has been made to ensure the accuracy, authenticity, and proper citation of the content published in this journal, neither the Editorial Board nor IJLRA shall be held liable or responsible, in any manner whatsoever, for any loss, damage, or consequence arising from the use, reliance upon, or interpretation of the information contained in this publication.

The content published herein is intended solely for academic and informational purposes and shall not be construed as legal advice or professional opinion.

**Copyright © International Journal for Legal Research & Analysis.
All rights reserved.**

ABOUT US

The *International Journal for Legal Research & Analysis (IJLRA)* (ISSN: 2582-6433) is a peer-reviewed, academic, online journal published on a monthly basis. The journal aims to provide a comprehensive and interactive platform for the publication of original and high-quality legal research.

IJLRA publishes Short Articles, Long Articles, Research Papers, Case Comments, Book Reviews, Essays, and interdisciplinary studies in the field of law and allied disciplines. The journal seeks to promote critical analysis and informed discourse on contemporary legal, social, and policy issues.

The primary objective of IJLRA is to enhance academic engagement and scholarly dialogue among law students, researchers, academicians, legal professionals, and members of the Bar and Bench. The journal endeavours to establish itself as a credible and widely cited academic publication through the publication of original, well-researched, and analytically sound contributions.

IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines. All manuscripts are subject to a rigorous peer-review process to ensure academic quality, originality, and relevance.

Through its publications, the *International Journal for Legal Research & Analysis* aspires to contribute meaningfully to legal scholarship and the development of law as an instrument of justice and social progress.

PUBLICATION ETHICS, COPYRIGHT & AUTHOR RESPONSIBILITY STATEMENT

The *International Journal for Legal Research and Analysis (IJLRA)* is committed to upholding the highest standards of publication ethics and academic integrity. All manuscripts submitted to the journal must be original, unpublished, and free from plagiarism, data fabrication, falsification, or any form of unethical research or publication practice. Authors are solely responsible for the accuracy, originality, legality, and ethical compliance of their work and must ensure that all sources are properly cited and that necessary permissions for any third-party copyrighted material have been duly obtained prior to submission. Copyright in all published articles vests with IJLRA, unless otherwise expressly stated, and authors grant the journal the irrevocable right to publish, reproduce, distribute, and archive their work in print and electronic formats. The views and opinions expressed in the articles are those of the authors alone and do not reflect the views of the Editors, Editorial Board, Reviewers, or Publisher. IJLRA shall not be liable for any loss, damage, claim, or legal consequence arising from the use, reliance upon, or interpretation of the content published. By submitting a manuscript, the author(s) agree to fully indemnify and hold harmless the journal, its Editor-in-Chief, Editors, Editorial Board, Reviewers, Advisors, Publisher, and Management against any claims, liabilities, or legal proceedings arising out of plagiarism, copyright infringement, defamation, breach of confidentiality, or violation of third-party rights. The journal reserves the absolute right to reject, withdraw, retract, or remove any manuscript or published article in case of ethical or legal violations, without incurring any liability.

AI AND COPYRIGHT LAWS IN INDIA- QUESTION OF ORIGINALITY AND OWNERSHIP

AUTHORED BY- YASHASVINI

Student at Final year BA LLB

Shri Navalmal Firodia Law College

Affiliated by Savitribai Phule Pune University

Abstract

The emergence of Artificial Intelligence as a technology to generate content has sparked various discussions related to novelty of creative work, creative process, data privacy, originality and ownership. This paper delves deep into the analysis and interpretation of Indian Copyright Laws in relation to literary, artistic and cinematographic work created by Artificial intelligence. It evaluates the originality of work created by AI on various tests developed through judicial pronouncements in case of derivative works. The paper comparatively studies the definition of authorship for work created by humans and work created by computer programmes according to the provisions of Indian Copyright Act, 1957.

It delves deep into the concerns of copyright infringements of the input data used to train AI with the help of recent case laws on AI companies like Open AI and Stability AI. This paper delves deep into the Analysis and Interpretation of Indian Copyright laws in work created by artificial intelligence. It evaluates the Originality of work created by AI on various tests developed through judicial pronouncements in case of derivative works. It comparatively studies the definition of Authorship for work created by humans and work created by computer programmes according to the provisions of Indian Copyright Act, 1957. This paper also gives a useful understanding of copyright licensing and its limitations. It clearly demarcates the boundary of fair use and identifies the extent of breach of these boundaries by AI companies. It also compares the stance of US law with Indian Copyright laws and suggests models that can help make a collaborative environment between AI companies and the companies they derive data from.

Introduction

Today, Generative AI is being extensively used to dish out creative content on the command of a single prompt. The content is available in the form of videos, images, or even literary works. To create this content it uses databases of a number of varied data. Artificial intelligence through algorithms and rules identifies and mimics patterns, eliminates things that are not required, and arranges them in recognized format. This process requires a large database to identify commonalities and make the content more relevant and accurate. This data may come from different sources like internal company data which includes CRM (Customer Relationship Management System) consisting of customer's feedbacks and activities on the platform and ERPs. A large volume of this data comes from public databases, social media platforms, and third-party sources. Sometimes the company may not procure this data by ethical means. Furthermore, even if the company/business does hold a copyright license indicating procurement of data by ethical means, the use of the data might still be well beyond the limited scope of the license granted. Many leading generative AI companies are broiling under various lawsuits regarding copyright infringement for the unauthorized use of data in creating derivative work.

AI is a computer technology that does machine learning which means it evolves its learning and intelligence artificially like its name suggests, with the help of a collection of data in the form of images, literary works, videos, audio, computer programmes, algorithms that are feeded into it. Artificial Intelligence analyses this data, identifies a common pattern and arranges it into appropriate pattern to come up with the output which can be either the solution to a problem or something entirely new.

Today, AI can be used in various creative as well as computational pursuits. It can be used to create a picture, image, videos, written work. It can also be used to create a computer programmes or codes to either develop a software application or a website.

All these functions requires a large amount of data and the question that arises is where does this data come from and whether or not the data is procured by ethical means. This is where the question of copyright takes the center position.

The copyright laws in India presently do not provide the answers pertaining to the ownership of content created by AI or regarding the legitimacy of the ownership of data used to create

such content.

This paper discusses the stance of Indian Copyright Laws with respect to ownership of content created and the data used to create such content in this research Article.

Methodology

The article adopts a qualitative doctrinal legal research methodology, it uses various secondary sources of information like articles, journals, books, statutes, legislations and judicial pronouncements and legal commentaries to identify the issues underlying the use of Gen AI. It analyses the Indian law and case laws to find out the current stance of Indian Copyright laws on work generated by AI. This methodology employs the method of collection of information from various sources and use these materials to interpret the legal question and develop a coherent response based on critical reasoning.

Analysis of the Indian Copyright Act, 1957 with respect to Generative AI Originality- Copyrightability of work created by AI

The first question which arises when assigning copyright to a work is whether the work is original or not as Section 13 of Indian Copyright Act, 1957 states that for copyright to subsist in a work, it needs to be original literary, dramatic, musical and artistic work.

The concept of originality is nowhere explained or defined in the Act.

This is why there are various tests evolved by judicial pronouncements to determine originality of a work.

Firstly, it should owe its origin to the author. There are various instances where even if the work is quite similar to other, if it is proved to be created independently of other work then it can be copyrightable.

Work created by AI fails on this test of originality as the work created is very much dependent on the data used.

In case of *Rupendra Kashyap v. Jiwan Publishing Houseⁱ*, the court stated that the word

originality did not imply originality of ideas but merely that the work in question shouldn't copy the other work.

In case of **derivative works** like abridgment, adaptations and translations - the skills and labour employed in selecting and arranging the subject matter gives rise to an original literary work. This is called the "sweat of the brow test".

In the case of *Eastern Book Company v. D.B. Modak*ⁱⁱ however "minimum of creativity test" was introduced where the work need not be entirely novel but it should have a flavour of creativity.

The two components of the "minimum creativity test" is: originality and skill and judgement. In AI generated words when a person simply gives a prompt based on which the work is created, the work lacks skill and judgement of the person giving the prompt as it is primarily the AI algorithm responsible for generation of that work.

Secondly, to determine the originality of AI work, the question which needs to be answered is *whether Databases can be considered as original work?*

Databases were included in the definition of "literary work" in section 2(o) by an amendment in 1999.

The explanation clause in Section 43 of the Information Technology Act, 2000 defines the term computer database as- "a representation of information, knowledge, facts, concepts or instructions in text, image, audio, video that are being prepared or have been prepared in a formalized manner and have been produced by a computer, computer system or computer network and are intended for use in a computer, computer system or computer network."

*Burlington Home Shopping Pvt. Ltd. v. Rajnish Chibber*ⁱⁱⁱ was a case where a mail order service sued a former employee for copyright infringement. The plaintiff had developed a database of customer information over three years. The defendant, Rajneesh, had obtained a copy of this database and used it for his own purposes.

The court found that the defendant had imitated the plaintiff's database, copying even minor details. This was considered a clear case of copyright infringement. The court held that the

plaintiff would suffer irreparable harm if the defendant were allowed to continue using the database.

Thus, if the work passes the minimum creativity test it can be designated an original work.

Authorship- Who owns the work created by Gen AI

The next issue concerning work created by Generative AI is with respect to ownership rights.

Once the minimal creativity threshold is established for a AI generated work, it falls under the definition of original work and therefore become copyrightable, then the issue of ownership of such work is raised.

Who owns the rights of distribution, ownership, reproduction, broadcasting etc of the content? Whether it is the platform that provides the AI service or the person giving the prompt or the owners of the data that is used to generate that work?

Author and Owner according the Copyright Act, 1957-

Author is the creator of a work and owner is someone who has economic rights on the intellectual property. The fundamental rule is that author is the first owner of the work. Therefore it becomes crucial to identify the author of a work.

Section 2(d) of the Indian Copyright Act 1957 defines the word author in relation to different types of work.

Section 2(d)(vi) of the Copyright Act states the meaning of author *in relation to any literary, dramatic, musical or artistic work which is computer generated, the person who causes the work to be created.*

Where ordinarily person who creates the work is the author, in case of computer generated works author is the person who 'causes' the work to be generated. But since Gen AI uses training data to produce work, it becomes difficult to identify the author of the work.

The Act however does not define computer generated work. It is crucial to define computer generated work to see if AI comes under this category. If yes, then the person who gives out the prompt becomes the author of the work as he causes the work to be created. But since the

developer of the programme as well as the author of input data has also caused the work to be created. It again traces back to the same ambiguity.

It is unclear if the remoteness of the cause of the creation should be taken into account or the contribution of the respective stages in the content. It is also unclear how to find out and compare the contribution of different facets at different stages.

To solve this problem it is necessary for a clear interpretation of this statute through judicial pronouncements or fresh legislation regulating AI and Copyright laws.

RAGHAV AI^{iv} painting software was the first AI tool recognized by Indian copyright office. When the co-author of the software commissioned his painting, the notice of withdrawal was sent by Copyright office citing the definition of author in relation to an artistic work under Section 2(d) (iii) which is supposed to be the artist who creates that work.

This case was later taken up in Canadian and US jurisdiction for protection where Canada granted the protection US rejected it as US mandated for a work to be copyrightable to be created by a human being. India also seems to be taking a similar stance to US.

This becomes even more clear by the case of *Rupendra Kashyap vs Jivan Publishing* ^v Delhi High court held that the questions papers under the head of CBSE compiled by a natural person which are the paper setters and not an artificial person which the CBSE board itself.

"By virtue of Section 17 of the Copyright Act copyright vests in the author of the work. The question paper is a "work" within the meaning of the Copyright Act. The author of the examination paper is a person who has compiled the questions. The persons who does this compiling, is a natural person, a human being, and not an artificial person. The Central Board of Secondary Education is not a natural person. The Central Board of Secondary Education would be entitled to claim copyright in examination papers only if it establishes that natural person have been instructed to compile the question papers."

In the case of *Tech Plus Media Private Ltd vs Jyoti Janda & Ors*^{vi}, the court stated that a juristic person since it is incapable of holding property cannot be the author of a copyrighted work unless there is an agreement between the juristic person and the author of the said work.

Input Data- Fair use of Data and Scope of Copyright Licensing

Generative AI is feeded large amounts of data to train the software with relevant data. It has become even more important to keep large database with the advent of LLMs (Large language models) for Text based AI like Chat GPT and Open AI to build their natural language understanding (NLU) and natural language processing (NLP) capabilities. There are unsupervised machine learning algorithms that are feeded for the AI to identify patterns using neural networks to form structure to train itself.

This entire training process requires large amounts of data which are sourced from company's own CRMs (Customer Relationship Management) and ERPs (Enterprise Resource Planning), Social Media platforms and other public databases which sometimes may contain pirated copies of the said data.

The present day practice is such that vast amount of data is obtained by statutory licensing for which it is necessary to identify the scope of copyright licensing and fair use of copyright. AI companies use TDM (text and data mining) technology to train AI models to analyze patterns in copyright protected databases. The extent of use of this TDM technology needs legal consideration.

In US, the data collection for training or TDMs has been declared as fair use. It has released limitations on TDMs under which collection of data mining is legally authorised. This comes as a great relief to new emergents in the business of Gen AI.

The Indian Copyright Act, 1957 also provides for fair usage of work protected from Copyright infringement under Section 52 (1) (a) The act explains that mere storage of information for research purposes, criticism, review and reporting of current events or reporting of lectures shall not constitute infringement of copyright under the act.

There are various factors which should be considered when determining if the usage comes under fair use. These are-

- If the work is of commercial nature or for non profit educational purposes.
- the nature of copyrighted work
- the amount of work and importance of the piece of work used in creating derivative work
- the potential risk to undervalue the market value of the copyrighted work.

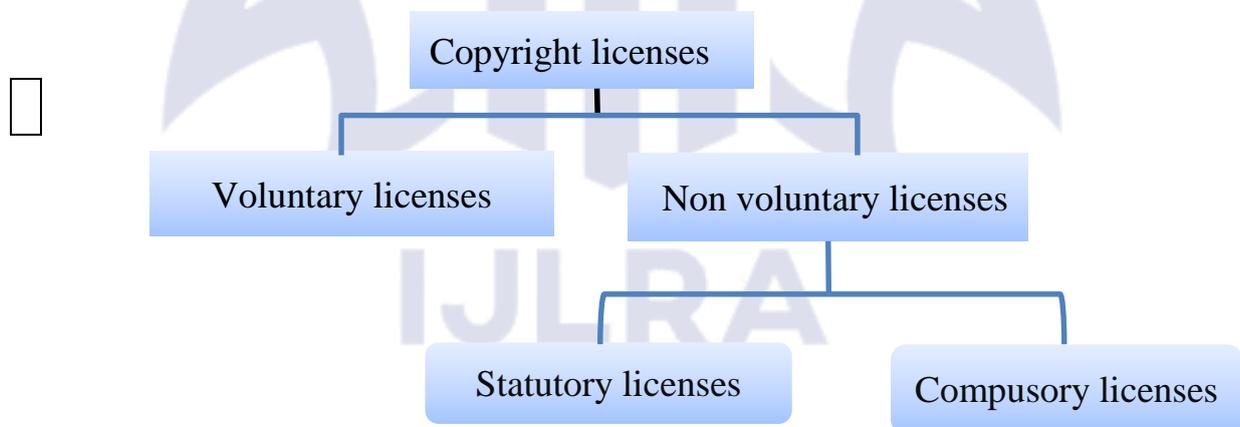
When we apply this provision to the use of data in training Gen AI, we observe that data is stored for machine learning, but unlike humans, machine learning works in a different way, it requires copying of the data to train artificial intelligence of the software. Hence, it cannot be regarded as fair use of copyrighted material.

The Berne Convention also realises the importance of balancing the society's interest as a whole as well as the rights and interests of the owners.

This is why it has provisions for nations to permit reproduction of copyrighted works in certain cases, provided that it does not facilitate exploitation of the work and it does not create unreasonable prejudice against owners interest.

Section 52 (1) (f) provides certain permitted reproductions of copyrighted work in the form certified copies published by due authorization of law for the permitted time limit.

Scope of Copyright Licences



Copyright licences are of two kinds- Voluntary and Non- voluntary

Non voluntary licences are further divided into two categories - Statutory licenses and compulsory licenses

Compulsory licenses are those where the copyright owner and the buyer enters into an agreement for authorisation of the work. The terms of authorization and prescribed conditions can be negotiated.

Statutory licenses which is commonly granted to AI companies are the licenses where Royalty to owner is fixed and the work can freely be used as long as the fixed fees is being paid to the designated body.

Section 31D inserted by 2012 Amendment Act provides the conditions that are required to be fulfilled when obtaining a statutory license.

-Prior Notice of duration of use, extent of use and territorial coverage of use to should be given to the owner.

-Fixed Royalty payment to designated body

-Legitimate credits given to the authors and performers of work

-No unreasonable alterations

-Maintenance of records and books of account open to inspection by owner.

Since AI modified and alters the data to generate creative work on its own, it breaches the condition of fair use even when statutory license is granted. It also lacks in giving due credits to the original author of the work.

Due to these serious infringements there are various lawsuits on AI based companies

In *Tremblay v. Open AI Inc*^{vii}, Open has AI has been accused of allegedly copying literary works.

In *Getty Images v. Stability AI*^{viii}

Where the Stability AI has been found to be involved in unauthorised scraping. Scraping means theft of personal data without consent.

This case involved theft of trademark of the image sourcing company called Getty Images. The AI has been used to generate the logo of Getty images with minor calculated alterations which is not only the infringement of copyright but also promoting concealment and unauthorised use of trademark.

These issues suggests that the AI companies should take accountability of the fair use of data which is valuable to their performance.

Conclusion

AI will continue to advance furthermore in coming years and the myriad of issues relating to copyright of input data will stand in the way of technological development of AI in India. The lack of laws protecting AI generated works from copyright infringement will destroy the incentive of developers. Therefore it is crucial to draft a legislative framework for Copyright Laws specifically for AI related content which assigns exact ownership of such works, revises the limitation of copyright licences, drafts provisions to sufficiently credit and compensate the original author.

AI companies should also be accountable for fair use of the data and create revenue sharing models for companies sharing their data with them with due payments of Royalty to them. Policies should be made encouraging public databases for training AI.

ⁱ 1996(38)DRJ81

ⁱⁱ 2008 AIR SCW 49

ⁱⁱⁱ (1996)PTR 40 (Del.)

^{iv} <https://spicyip.com/2023/12/ankit-sahnis-ai-co-authored-artwork-denied-registration-by-us-continues-to-be-registered-in-india.html>

^v 1994 IAD Delhi 1

^{vi} 2014 (60) PTC 121 (Del)

^{vii} 3:23-cv-03223-AMO

^{viii} 2023 ewhc 3090 (ch)