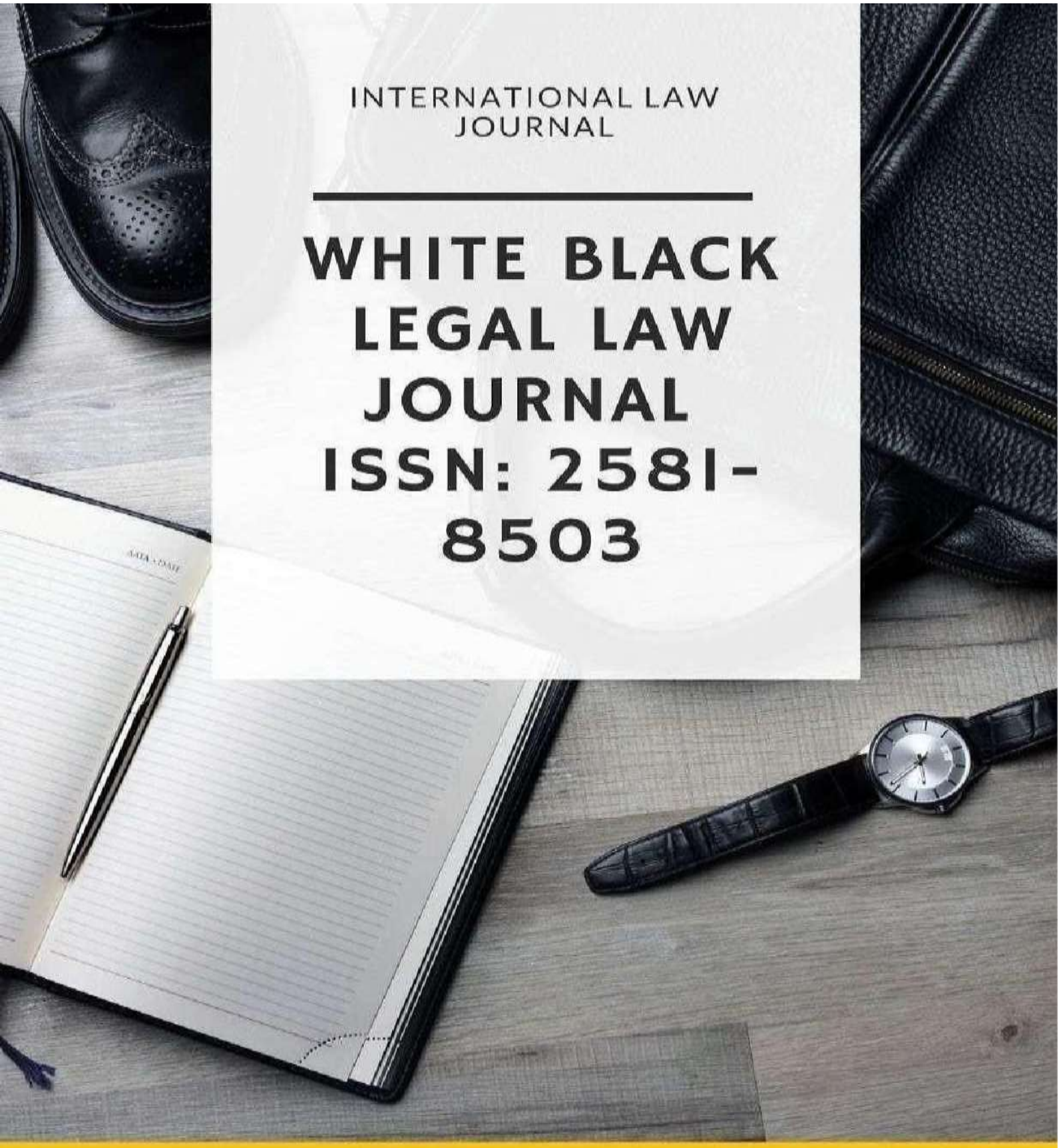

INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL

ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

No part of this publication may be reproduced or copied in any form by any means without prior written permission of Editor-in-chief of White Black Legal – The Law Journal. The Editorial Team of White Black Legal holds the copyright to all articles contributed to this publication. The views expressed in this publication are purely personal opinions of the authors and do not reflect the views of the Editorial Team of White Black Legal. Though all efforts are made to ensure the accuracy and correctness of the information published, White Black Legal shall not be responsible for any errors caused due to oversight or otherwise.

EDITORIAL TEAM

Raju Narayana Swamy (IAS) Indian Administrative Service officer

Dr. Raju Narayana Swamy popularly known as Kerala's Anti-Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala. He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University. He also has an LLM (Pro) (with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and a professional diploma in Public Procurement from the World Bank.

Dr. R. K. Upadhyay

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB, LLM degrees from Banaras Hindu University & PHD from university of Kota. He has succesfully completed UGC sponsored M.R.P for the work in the Ares of the various prisoners reforms in the state of the Rajasthan.

Senior Editor

Dr. Neha Mishra

Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; PH.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St. Louis, 2015.

Ms. Sumiti Ahuja

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,
Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing PH.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.

Dr. Navtika Singh Nautiyal

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in 'Intercountry adoption laws from Uttranchal University, Dehradun' and LLM from Indian Law Institute, New Delhi.

Dr. Rinu Saraswat

Associate Professor at School of Law, Apex University, Jaipur, M.A, LL.M, PH.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

Dr. Nitesh Saraswat

E.MBA, LL.M, PH.D, PGDSAPM
Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.
More than 25 Publications in renowned National and International Journals and has authored a Text book on CR.P.C and Juvenile Delinquency law.

Subhrajit Chanda

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); PH.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

ABOUT US

WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal provide dedicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

"TEXT AND DATA MINING, GENERATIVE AI AND SECTION 52: RE-EXAMINING INDIA'S COPYRIGHT FRAMEWORK"

AUTHORED BY - RAIBA JESTUS

Abstract

These days, AI systems are capable of creating artistic works like music, poetry, and paintings. Artificial intelligence (AI) is "trained" using text and data mining (TDM) approaches, as well as for other AI research and development. The governance of TDM in India is the main topic of this article. According to the author, copyright issues pertaining to TDM activities may be covered by Section 52 of the Indian Copyright Act. Nonetheless, legislative action is required to ensure equitable governance and address certain related issues. The article's first section discusses the extent to which copyright issues related to TDM in India are governed by the "fair dealing" principle. The author argues that India's socioeconomic circumstances justify a permissive exemption of copyright to carry out TDM operations in the second section, which also explains the TDM exceptions in other countries. Additionally, the author contends that a limited exception, like the EU's Directive on Copyright in the Digital Single Market (CDSM), would be harmful in general progress. The author has asserted that India might benefit from permissive TDM exclusions similar to those in Japan, despite efforts by AI companies to abuse copyright holders. The final portion uses significant socioeconomic factors to bolster the liberal TDM exceptions thesis, including the state of the home economy and linguistic, social, and cultural diversity.

Keywords: AI, Text and Data Mining, Copyright, IPR, Data

I. Introduction

Protecting the rights of the copyright owner of the data used for TDM is essential while developing TDM and developing policy to handle its copyright issues. But it's also critical to take into account how India's legal system regulates TDM. A key component of AI policy is how AI will affect society. For this reason, nations are giving it careful thought while

developing their AI policy plans[1]. The creator of the artwork receives the intangible copyrights as compensation. However, when it comes to copyrights, the rights of society as a whole have always been given equal weight. Section 52 of the Indian Copyright Act establishes the fair dealing principle in India. This raises the question of whether India can impose a general set of regulations on all types of TDM, including text, images, and videos. The Indian Copyright Act is well-suited to safeguard AI-generated content, according to the Press Information Bureau (PIB)[2]. The TDM and copyright concerns are not discussed. However, according to Section 52 of the Act, if a user of Generative AI wishes to utilize their creations for commercial purposes, they must get permission from the copyright owner.

This article makes the case that the current system is inadequate to address the copyright concerns related to TDM. The implications of adopting the TDM policies of the EU and Japan regimes in India are also discussed in this article. The paper proposes a policy akin to the Japanese AI controlling TDM policy, with certain unique obstacles, based on socio-economic conditions such as India's cultural, linguistic, and economic uniqueness. Certain laws address IP-related difficulties and the administration of AI generation subjects[3]. The TDM and copyright concerns around the use of data for AI training, however, are not discussed.

II. The Current Legal Position of TDM in India

Two important legislation that protect personal data are the IT Act[4] and the SPDI Rules[5]. The TDM and other AI-related issues are not included in any of these legislation. A framework for security, data privacy, and an accountability system is needed in the Indian policy arena. Copyright concerns regarding. TDM is necessary to prevent problems like bias and discriminating generative AI output. Therefore, TDM needs to be addressed in order to address the issues raised in the advice. In the absence of regulation, MNCs might adopt other strategies

[1] Select Committee on Artificial intelligence of the National Science and Technology Council, '*National Artificial Intelligence Research and Development Strategic Plan' (2023), Inclusion of Strategies 3 on understand and address the ethical, legal, and societal implications of AI and Strategy 5 on develop shared public datasets and environments for AI training and testing.*
[2] Mossoff A, Saving Locke from Marx: The labour theory of value in intellectual property theory, *Social Philosophy and Policy,* 29 (2012) 283, https://ssrn.com/abstract=1983614
(accessed on 7 August 2025).
[3] The Copyright Act, 1957, Section 2 (d) (vi).
[4] Information Technology Act, 2000.
[5] Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011.

to control this matter, which would make regulation and governance more difficult[6].

The Fair Dealing Principle to Regulate TDM

Section 52[7] of the Copyright Act is where the idea of fair dealing originated in India. The Principle to create harmony or balance between the author's conflicting monopolistic interests and the creative goals of society as a whole, makes fair dealing necessary. What does TDM activity include, and does it even come under copyright protection? However, the fair dealing concept will apply in certain situations as specified by Section 52 of the Act. However, Indian courts have given the fair dealing concept an extremely wide interpretation, and it is reasonable to assume that it may also partially address the copyright concerns related to TDM.

Perspectives of the Fair Dealing Principle and Section 52

TDM via the prisms of the Fair Dealing Principle and Section 52

In India, the Court used fair dealing under Section 52 of the Act by looking at elements including "amount and substantiality of the portion used" and "purpose and character of the use"[8]:. Generally speaking, the fundamental idea that courts take into account is whether the goal of the latter work and the earlier effort is the same or significantly different. There won't be fair dealing if it is different[9]. It is impossible to define what constitutes "fair dealing," as Lord Denning once stated. It must be a matter of degree[10]. The courts must determine what constitutes "fair dealing" on a case-by-case basis due to the absence of a statutory definition.

Understanding Section 52 of the Act

Additionally, a number of decisions indicate that the legislative goal of Section 52 of the Act is to protect the rights of Indian society as a whole. The Court declared in the Wiley Eastern Ltd. Case that Section 52 of the Act serves to safeguard the rights outlined in Article 19(1) of the Indian Constitution. The purpose of this section is to encourage private study, inquiry, criticism, and review[11]. The 2012 change to Section 52 of the Copyright Act demonstrates how broad the scope is and how it can be expanded in accordance with societal demands. Indian

[6] Dermawan A & Mezei P, Artificial intelligence and consensus-based remuneration regime in Southeast Asia, 12, https://ssrn.com/abstract=4625850 or http://dx.doi.org/ 10.2139/ssrn.4625850 (accessed on 12August 2025).
[7] The Copyright Act, 1957, Section 52.
[8] *Civic Chandran* v *C Ammini Amma*, [1996] PTC 16 670.
[9] *ESPN Stars Sports* v *Global Broadcast News Ltd and Ors*, 2008 (36) PTC 492 [Del] [Para 17].
[10] *Hubbard* v *Vosper*, CA 1971 [1972] 2 WLR 389.
[11] *Wiley Eastern Ltd & Ors.* V *Indian Institute of Management*, 61 [1996] DLT 281, At Para 19 Court said that "The basic purpose of Section 52 is to protect the freedom of expression under Article 19(1) of the Constitution of India- so that research, private study, criticism or review or reporting of current events could be protected"

courts have frequently cited US judgments and taken the 4-step test into consideration[12].The fair dealing case was adjudicated by Indian courts based on the specifics of the issue[13]. However, in contrast to the US, Section 52 of the Act specifies the circumstances in which fair dealing will be relevant. However, the court gave it a fairly liberal interpretation. The fair dealing principle may apply to TDM because it is not specific. However, since copyright is just one part of the problem, the TDM cannot be guided only by the fair dealing principle. The concept of "fair dealing" is multifaceted rather than one-dimensional, and it should be considered from the perspectives of the public, the owner, and the writers[14]. This broad approach makes it possible for Section 52 of the Act to apply to the TDM.

India has an extremely high rate of copyright infringement when compared to other nations. India is ranked 43rd out of 53 nations in terms of IP rights enforcement, including infringement, the civil and criminal legal channels accessible to copyright holders, and the power of customs authorities to conduct border controls and inspections[15]. This shows that copyrighted content is readily available. Furthermore, in certain instances, it was decided that even if the user is aware that the source is an infringed copy, using it in a way that makes sense and qualifies as "fair use" does not constitute bad faith. It would be deemed fair use if the infringed subject was used for a transformative purpose[16].The developers might try to take advantage of this to get around the copyright owner's consent, which could lead to problems. This could encourage the country's unlawful use, sale, and infringement of copyrighted content. In this case, the crucial queries to pose are: What is the current state of the copyright holder's exclusive economic rights being used for non-commercial purposes? How do you feel about data being temporarily stored? Which works would be covered by Section 52 of the Act's fair dealing doctrine and which would not?

[12] In the case, *The Chancellor University of Oxford* v *Narendra Publishing House & Ors*, CS (OS) 1656/2005,the Delhi High Court referred to the case *Harper & Row* v *Nation Enterprises* 471 US 539 [1985] and decided on the basis of four factor test that the amount and value of the content taken from the original work would be considered to decide its effect on the market share of the original work.

[13] In *Super Cassettes Industries* v *Mr Chintamani Rao & Ors.* on 11 November 2011; while citing the American Case *Harper & Row Publishers, Inc.* v *Nation Enter.*, 471 U.S. 539, 588, 85 L.Ed 2d 588, 105 S. Ct. 2218 [198] Delhi High Court held that work used for criticism and review about the theme of original work is fair dealing as it does not harm the future market of the film.

[14] Justice Singh P M, Evolution of Copyright Law: The Indian journey, *Indian Journal of Law and Technology*, 16 (2) (2020) 54.

[15] US Chamber of Commerce, Art of the impossible, US Chamber of international IP indexglobal innovation policy centre, 2020, 67.

[16] Carroll M W, Copyright and the progress of science: Why text and data mining is lawful, *University of California, Davis*, 53 (893) 956.

According to Section 2(m) of the Act, reproducing a dramatic, literary, musical, or creative work constitutes an infringement if the reproduction violates the Act's provisions[17] According to Section 52, if copyrighted material is utilized for personal use, including research, it does not constitute infringement to save any work in any electronic medium, including the incidental storage of any computer program that is not an infringing copy under the Act. Thus, it would not apply if such material were used for commercial purposes.

Furthermore, Clauses (b) and (c) of the Act do not prohibit the incidental or temporary preservation of work or performance for any technological electronic transmission or communication process to the public. Therefore, it is not an infringement to temporarily store the data in order to carry out the action allowed by Section 52 of the Act.As a result, using the copyrighted work for non-commercial purposes that are covered by Section 52 of the Act might not be considered an infringement.

The final question's response would vary from instance to case. In the Authors Guide v. Google case, the Court determined that Google Books sought to provide consumers with important information about the books based on the "need" for the infringement and "literal necessary" to achieve the objective[18]. The Court relied on the reason for retaining copies of the protected content in A.V. v. iParadigms. It is not an infringement if the goal is legitimate and required to complete the operation[19]. We can argue that if a disagreement emerges in the TDM case, the courts might adopt comparable strategies.

Since TDM action is a non-expressive use of the copyrighted material, the concept behind the protected work should not be covered by the protection because it is a well-established principle in India that neither ideas nor expressions are protected by copyright. The High Court ruled in the Kartar Singh Giani case that "there are two points connected with the meaning of the expression "fair," in fair dealing." "An intention to compete and to derive profit from such competition" comes first, followed by "motive of the infringer."[20]. The TDM example lacks both of these components. Therefore, if TDM is used without permission in identical

[17] Additionally, it is also infringement, "in relation to a cinematographic film, a copy of the film made on any medium by any means; and, in relation to a sound recording, any other recording embodying the same sound recording, made by any means"
[18] *Authors Guild* v *Google* 721 F. 3d 132 [2nd Cir. 2015].
[19] *A.V. ex rel. Vanderhye* v *iParadigms*, L.L.C., 562 F.3d 630 [4th Cir. 2009]
[20] *Kartar Singh Giani* v *Ladha Singh*, [1934] SCC Online Lah 277

circumstances, it might not result in copyright infringement. The AI sector is in a stage where these issues must be resolved for TDM to be possible without any problems, and the Court may take years to resolve them. Therefore, legislative or executive action is necessary for all of these reasons.

III TDM in different Legal Jurisdictions

The European Directive on Copyright in the Digital Age is the EU statute that addresses this matter.

Single Market (2019/790). The required exemption under EU copyright law, which exempts the copying of content protected by copyright, is discussed in Article 3 of the directive. Additionally, subject content protected by the sui generis database, SDGT, is exempt from extraction under this article.

Similarly, all beneficiaries have access to Article 4. Beneficiaries have access to any kind of data, with the exception of situations in which the copyright owner has chosen to forego this exception, as Article 4 grants the copyright holder this privilege[21].

According to the US Copyright Office's most recent "Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence," the applicant must include a brief description of their contributions to the work along with a disclosure of the use of AI-generated content. TDM exploitation cannot be considered an infringement since it is not an act of enjoyment. Even with such limited exceptions, a number of problems could make it difficult to execute the EU exceptions[22].

Only research institutions and cultural heritage for scientific research are eligible for Article 3 exceptions in the EU. Additionally, Article 4 allows the right holder to choose not to take advantage of this article's exception. As a result, the academics contend that the exceptions are insufficiently broad to permit R&D in the area of AI. According to academics, even though the copyright exception under these laws is supported by strategic policy objectives, it may be

[21] Directive of the European Parliament and of The Council of [2019] (EU) 2019/790.
[22] Dermawan A, Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?, *Journal of World Intellect Property*, 15 (2023) 1

conceptually faulty[23]. According to the statistics, the EU is falling behind the US and Asia (particularly China) in terms of TDM patents. The three main reasons for this are legal ambiguity on how TDM activities are handled in the EU and its copyright law, a lack of knowledge and expertise, and infrastructure difficulties[24]. This could maintain the expense of developing AI and serve as a barrier.

The Japanese Copyright Act, which went into effect on January 1, 2019, was modified by the Japanese government on May 18, 2018[25]. In contrast to the EU, Japan has more permissive exceptions, and the Japanese government has made sure that copyright shouldn't impede the advancement of AI. TDM is allowed without prior agreement from the relevant right- holders unless it unreasonably jeopardizes the copyright holder's interests if the purpose of the exploitation is not enjoyed by the exploiter or making another person enjoy the work, unless the exploitation is detrimental to the copyright holder's interests."[26] The TDM exclusions in Japan are extremely expansive since they can be used for both commercial and non-commercial activities without the right holders' permission. The TDM exclusions in Japan are extremely expansive since they can be used for both commercial and non-commercial activities without the right holders' permission. To put it another way, exploitation is always acceptable. Unlike the EU, India chooses permissive TDM exceptions due to a number of socioeconomic concerns. The AI industry in India's domestic economy is still in its infancy, and the competitive market necessitates easy access to data for AI development; otherwise, it would act as a barrier and cause development to be delayed. Additionally, the application of AI in other sectors may have a spillover effect on other pertinent sectors. Additionally, if data availability is limited by stringent TDM restrictions, linguistic, social, and cultural diversity may provide significant issues for society.

[23] Margoni T & Kretschmer M, A Deeper Look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology, *GRUR International*, 71 (8) (20) (2022) 686
[24] The exception for text and data mining (TDM) in the proposed directive on copyright in the digital single market-technical aspects, European Parliament, https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf, 2.
[25] Japan amends its copyright legislation to meet future demands in AI and big data (European alliance for Research Excellence, 3 September 2018), 5 April 2023
[26] Dermawan A, Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes?, *Journal of World Intellect Property*, 11 (2023) 1

IV Significance of liberal TDM

Since AI is still in its infancy, developing precise long-term regulatory strategies is difficult. This makes regulating AI a challenging issue. It is difficult to predict how copyright law may react to work created by artificial intelligence. It is evident that in order to cover and address these challenges, the legal structure controlling copyright laws will undergo major alteration[27]. In its draft of the AI Act, the EU places a high priority on safeguarding its citizens.

Economic Aspects

The 161st Parliamentary Report recognizes that artificial intelligence (AI) and its applications to revenue generation and the Republic of India's overall economic health, along with its demonstrably positive impact on the country's technological innovation capabilities, require the controlled and secure expansion of AI within the country[28]. AI has the potential to boost the Indian economy by USD 957 billion by 2035, according to this analysis. Big companies like Meta, OpenAI, and Google already have a competitive edge over MSMEs in innovation and R&D in the field of artificial intelligence due to the data's accessibility[29]. These businesses would be able to maintain cheap production costs with easy extra data to train the AI model, which would disadvantage MSMEs[30]. Even though AI-based technologies present enormous benefits for developing nations, this new wave of technological advancement may also present considerable obstacles[31].

MNCs may relocate from developing to developed nations (for luxury goods) as a result of establishing their base close to consumers[32].

The nation must provide AI solutions in industries other than manufacturing (shipping,

[27] INDIAAI, Impact, opportunities and Challenges of generative AI, *A MEITY, NEGD & NASSCOM Initiative*, 38
[28] Parliament of India, *Review of the Intellectual Property Rights Regime in India* 161st Report of the Committee, Department Related Parliamentary Standing Committee on Commerce, 2021, https://www.tbalaw.in/_files/ugd/f4c767_53be272a8c0341faaa97f62c0e84617b.pdf, 31
[29] Ziaja G M, The text and data mining opt-out in Article 4(3) CDSMD: Adequate veto right for rightholders or a suffocating blanket for European Artificial Intelligence Innovations? *Journal of Intellectual Property Law & Practice*, 25 (2024) 3https://doi.org/10.1093/jiplp/jpae025
[30] Lucchi N, ChatGPT: A case study on copyright challenges for generative artificial intelligence systems, *European Journal of Risk Regulation*, 12 (2023) 1, doi:10.1017/err.2023.59
[31] Ernst E *et al.*, The economics of Artificial Intelligence: Implications for the future of work, ILO future of work, Research Paper Series, 5 (2018) 26
[32] Policy Department, Directorate-General for External Policies, Directorate General for External Policies of the Union Two briefings and an in-depth analysis on Data flows, artificial intelligence and international trade: impacts and prospects for the value chains of the future, PE 653.617 - November 2020, 9.

services, etc.) that would allow MNCs to maintain their manufacturing base in India in order to retain them in the domestic economy. However, MNCs would be able to relocate their businesses from advanced economies to nations with inexpensive skilled labor thanks to services that are becoming digitally transferable and highly labor-intensive, particularly the medium- to high-skilled labor services, AI, and digital technology[33]. The services industry is a significant part of the Indian economy, accounting for over half of the country's GDP[34]. In order to envision AI adoption in the Indian service sector, infrastructure, governance, and legislation must be developed. This would require a practical, low-cost method of creating AI solutions for the services industry.

Every sector needs to develop AI solutions that would offset the labor factor with other crucial elements like low-cost manufacturing, affordable mobility and transportation, and AI-enabled infrastructure that would keep MNCs drawn to Indian domestic sectors in spite of the labor factor. AI companies require the policy space to develop their products in order to create these answers. Only when the data is readily accessible, the solution is impartial and accurate, and the development cost is low enough to make these solutions competitive and reasonably priced can such a policy space be established.

In TDM, the type of data is very important. The AI output will make more accurate recommendations with more data. But for an AI system to produce reliable results, data variation is just as important as volume. Therefore, unfettered access to more data would encourage greater competition and lead to better economic benefits for the nation's consumers[35]. However, care must be exercised because easy access to data may potentially lead to increased data misuse. The liberal TDM exceptions for India are supported by all of these observations.

The Linguistic Aspect

Larger tech businesses typically train AI language models in popular languages like English. The term of reference is still far ahead of the AI language models. This requires India, a

[33] OECD *Initial Policy Considerations for Generative Artificial Intelligence*, OECD Artificial intelligence Paper series No-1, September 2023 at 21.
[34] Ministry of Finance, Services contributed over 50% to GDP, 2022, https://pib.gov.in/PressReleasePage.aspx?PRID=1793804 (accessed on 15 November 2023).
[35] Policy Department, Directorate-General for External Policies, Directorate General for External Policies of the Union PE 653.617 - November 2020, 7

linguistically varied nation, to advance in AI research and development in order to create its own AI language models that are trained in Indian languages. India has a wide variety of languages. Since each state has its own language, it is crucial to invest in language resources and data storage organizations. AI should be trained using data from these languages. According to the 2011 Indian government census, 96.71 percent of the nation's citizens speak 121 different languages[36]. Access to technology is already hampered by the languages in the various parts of the nation. For instance, although it supports many other languages, Amazon Polly does not support any Indian languages. Hindi, Tamil, and Telugu are the only three Indian languages supported by Microsoft Azure's 45 language support options. In Hindi alone, 48 languages are acknowledged by society. Because of this, speech in India is quite informal, which could reduce the accuracy of AI speech models' output[37]. Enough high-quality data is required for the development and implementation of voice-based AI systems[38]. An AI system trained on Indian data may face a number of difficulties as a result of these characteristics.

The AI model's output would also depend on the type and quantity of parameters utilized to train the language model. In order to provide appropriate and less biased results, India may need to adjust the parameters of AI training models in some specific areas in accordance with Indian social and cultural settings[39]. The availability of data on each region-specific topic would be restricted and could only be found from that region due to the fact that each region of the country speaks a distinct language. For the AI solution to function effectively in such a situation, all the data should be readily available for TDM.

Prejudice, False Information, and Biased Results

Generally speaking, there are three ways that AI bias might infiltrate an AI system: data set bias, model design bias, and ML self-prejudice in training[40]. The online "commons" could be

[36] Government of India 'Census of India 201, Paper 1 of 2018 Languages' (2018), https://censusindia.gov.in/nada/index.php/catalog/42458/download/46089/C- 16_25062018.pdf, 4

[37] Deutsche Gesellschaft fürInternationale Zusammenarbeit (GIZ) GmbH, A Study on Open Voice Data in Indian Languages Implemented, 2020, 12, 17 & 39, https://www.bmz-digital.global/wp-content/uploads/2022/08/Study-on-Open-Voice-Data-in-Indian-Languages_GIZ- BizAugmentor.pdf.

[38] Lucchi N, ChatGPT: A case study on copyright challenges for generative AI systems, *European Journal of Risk Regulation*, 1 (2023) 17, SSRN: https://ssrn.com/ abstract=4483390 or http://dx.doi.org/10.2139/ssrn.4483390 (accessed on 12 June 2023), Carroll M W, Copyright and the progress of science: Why text and data mining is lawful, *University of California, Davis*, 53 (893) 956

[39] OECD, AI Language Models: Technological, Socio- Economic and Policy Considerations, *OECD Digital Economy Paper*, April 2023, No. 352, 34

[40] Carter C, Artificial intelligence and the future of work: A transatlantic comparison on the regulation of algorithmic and automated decision-making in the workplace, *TTLF Working Papers No. 91, Stanford – ViennaTransatlantic Technology Law Forum*,2022, 9.

harmed by AI language models that were trained on poor quality data[41]. Six risk areas related to AI language models have been identified by DeepMind. The TDM is connected to three of these regions[42]. The data used in the TDM process would determine the risk related to areas like discrimination, false information, and human-computer interaction[43]. These topics have also been taken into account by nations when developing their AI governance policies[44].

The TDM is necessary for all seven of NITI Aayog's defined principles[45]. It is crucial to use and implement generative AI in India in a way that respects human rights in order to reduce the significant hazards connected with this technology. Unintended consequences including prejudice, false information, and malicious use must be addressed in order to mitigate these dangers, and the government should assist companies in doing so[46]. This is impossible without giving businesses a fair amount of latitude to engage in TDM activities in order to develop AI solutions. Human rights like as equality and discrimination are also affected by the TDM law and regulation. It is crucial to understand that when an organization with legal or constitutional responsibilities uses an AI system, it should be impartial[47]. In India, the right to information under Article 19 of the Constitution would require accurate information to be skewed and false[48]. Articles 15 and 21 of the Constitution guarantee rights against discrimination. As was previously mentioned, Article 19 of the constitution serves as the foundation for Section 52 of the Copyright Act, and the liberal TDM may contradict the legislator's meaning. Therefore, the fundamental rights guaranteed by the Constitution cannot

[41] AI-Language Models: Technological, Socio-Economic and Policy Considerations, *OECD Digital Economy Paper*, April 2023, No. 352, 38.

[42] DeepMind, "Ethical and social risks of harm from language models" arXiv, Cornel University,http://arxiv.org/pdf/ 2112.04359, Dermawan A, Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese "nonenjoyment" purposes? *J World Intellect Property*, 15 (2023) 1.

[43] OECD, Hello, World: Artificial intelligence and its use in the public sector, *OECD Working Papers on Public Governance No*. 36, 139.

[44] The White House, A blueprint for an AI bill of rights, making automated systems work for The American People, 2022, 6.

[45] NITI Aayog, Responsible AI, AI For All, Approach Document for India: Part 2 - Operationalizing Principles for Responsible AI August 2021 at 4; India Impact, Opportunity, and Challenges of Generative AI,https://indiaai.gov.in/research- reports/impact-opportunity-and-challenges-of-generative-ai, 43; COM Building Trust in Human-Centric Artificial Intelligence, 2019,168, April 2019. https://ec.europa.eu/ futurium/en/ai-alliance-consultation/guidelines#Top; Shekar K *et al*., Principles for enabling responsible AI Innovations in India: An ecosystem approach the dialogue, (2023) 32, https://indiaai.gov.in/research-reports/principles-for-enabling- responsible-ai-innovations-in- india-an-ecosystem-approach.

[46] The Federation of Indian Chambers of Commerce & Industry (FICCI) The Economic Impact of Generative AI: the future of work in India, June 2023 at 23; Shekar K *et al*., Principles for enabling responsible AI Innovations in India: An ecosystem approach the dialogue, (2023)15.

[47] NITI Ayog, *Approach Document for* India Part 1 – Principles for Responsible AI, Responsible AI#AIFORALLFebruary 2021 at 29.

[48] The Constitution of India, Article 15.

be superseded by intellectual property rights.

V Conclusion

Since data and information are known to be among the most crucial components of technological advancement, intellectual property should serve to advance rather than impede technological advancement[49]. Framing these exceptions narrowly is inappropriate and could have long-term effects. It might pose a problem for the nation's AI industry's growth. Data is the driving force behind the AI sector, which makes it extremely complicated. All industries should be consulted extensively before the law is created, and the policy should be approached broadly and over an extended period of time. An excessive number of laws could hinder the AI industry and have an impact on the entire local economy.

India must strike a balance between advancing AI research and development and safeguarding the rights of right holders. It could be difficult to fix the biased AI solution developed during the development phase. Due to the inconsistent nature of the data, it is broken down state-wide by factors such as language, community peculiarities, religious beliefs, and various casts, creeds, and customs. Gathering such data to teach AI solutions would be difficult. There is a greater likelihood that the AI system's output will be prejudiced, discriminatory, and erroneous if there is less or no data. This would violate international regulations, such as the intellectual property principles outlined in Articles 7 and 8 of the TRIPS agreement, as well as the fundamental rights of the Indian constitution. The goal of Section 52 and the fair dealing principle may be violated if TDM is tightly regulated. In any case, either a liberal copyright exception with less dangers and more opportunities and inclusive development, or a stringent copyright exception with proportionally more risks and problems, would lead to the development of generative AI. We have the option.

When working in the policy space for AI governance in the IP sphere, policymakers would be guided by the reasons presented in this article to take into account the difficulties and characteristics of Indian society. This essay urges the governments of India and other nations with comparable linguistic and socioeconomic variety to take into account local quirks before depending on national laws and practices.

[49] Article 7 of TRIPS Agreement.