Generative AI, Text & Data Mining and the Fair Dealing Doctrine: Examining the New Problem with the Old Regime

Kailash Chauhan[†]

Technical Research Fellow, Intellectual Property Division (IPD), New Delhi High Court, Delhi — 110 003, India

Nowadays, AI systems can produce creative output, such as paintings, poetry, music, etc. Text and Data Mining (TDM) techniques are used to 'train' Artificial Intelligence (AI) and for other R&D in AI. This article focuses on the governance of TDM in India. The author has argued that Section 52 of the Indian Copyright Act may handle copyright aspects of TDM activities. Still, there is a need for legislative intervention for fair governance and to take care of some challenges associated with it. The first part of the article talks about the scope of the principle of "fair dealing" in governing copyright aspects associated with TDM in India. In the second part, while explaining the TDM exceptions in other countries, the author has argued how the socio-economic condition in India favours a liberal exception of copyright to carry out TDM activities, how the narrow exception like in the EU's Directive on Copyright in the Digital Single Market (CDSM) may prejudice the society and overall development in India. The author has argued that liberal TDM exceptions like Japan would benefit India even though there are challenges by AI firms to exploit the copyrights of the holders. In the third part, the argument of liberal TDM exceptions is supported by major socioeconomic factors like the state of the domestic economy and linguistic, social and cultural diversity.

While working on TDM and framing policy to govern the copyright challenges of TDM, it is crucial to protect the rights of the Copyright holders of the data used for TDM. However, it is also important to consider how the legal framework governs the TDM in India. The societal implications of AI are an important aspect of AI Policy. This is why countries are considering it seriously in their AI Policy strategies.[1] The copyrights are intangible in nature and are given as a reward to the creator of that art.[2] However, the rights of society at large have always been given equal importance when considering copyrights.[3] The fair dealing doctrine in India is enshrined in Section 52 of the Indian Copyright Act. Here, the question arises can we govern all forms of TDM, that is, videos, pictures, and literature, under a blanket set of rules in India? According to the Press Information Bureau (PIB), the Indian Copyright Act is well-equipped to protect the work generated by AI.[4] It does not talk about the TDM and copyright issues. However, it mandates the user of Generative AI to obtain permission from the copyright owner if they want to use their works for commercial purposes unless their use is exempted by fair dealing under Section 52 of the Act.

This article argues that the existing regime is insufficient to handle the copyright issues associated with TDM. This article also argues the implication of the TDM policies of the EU and Japan regimes if adopted in India. Based on socio-economic conditions such as India's cultural, linguistic, and economic peculiarities, the article suggests a policy similar to the Japanese AI governing TDM policy with some specific challenges. Some laws indicate the governance of AI generation subjects and issues related to IP.[5] However, there is almost no discussion of the TDM and copyright issues in the use of data in the training of AI.

Another regulation that deals with the governance of AI in India is the Ministry of Electronics and Information Technology (MeitY) advisory, which talks about unlawful content and biased or discriminatory information. However, the advisory may not be an appropriate solution. Para 2 (a) provides that the intermediaries and platforms who use AI/LLMs/generative AI, algorithms or software have to follow rule 3 (1) (b) of the IT Rules and provisions of IT Act 2000.[6] Under Section 3 (1) (b)

—————
[†]Email: Kailashsepato@gmail.com

(iv) of the said rules, compliance with IPR is also mandatory.[7] The compliance may not apply on the TDM and not even post infringement of LLM as it regulates the wrong done by the user, and the intermediary is obliged to intimate the same. Para 2 (b) of the advisory prohibits discrimination, bias or interference with the integrity of the electoral process. The biases and discriminatory output of the LLM depend on the TDM. Hence, to ensure this compliance w.r.t TDM should be considered. Similarly, under Para 2(c) of the advisory, the labelling of the unreliable or fallible output would also depend on how the LLM is trained. There are many cases where the courts have given immense importance to society's welfare.[8] So, following this precedent, it is important to give special consideration to society's welfare while considering the governance of TDM in India.

AI's huge economic potential in India needs to be kept in mind while considering the governance of Gen AI. AI is predicted to boost India's yearly growth rate by 1.3 percentage points. AI has the potential to account for US $ 957 billion, or 15% of the present gross value added, to India's GDP in 2035.[9] Hence, while working on the governance of copyright issues, this economic potential needs to be kept in mind.

State of Play of AI Governance in India

The IT Act[10] and SPDI Rules are major laws protecting personal information.[11] None of these laws talks about the TDM and other AI-related issues. The Indian policy space needs a framework that would create a framework for security, data privacy and a system that deploys accountability. The above mentioned Press Information Bureau publication does not cover copyright issues w.r.t. TDM. The 2024 MeitY advisory also does not cover the copyright issues associated with TDM and AI. Para 2 (b) of the advisory talks about maintaining the integrity of the electoral process. Issues such as bias, and discriminatory output of the generative AI, TDM is a prerequisite for the prevention of these issues. So, TDM must be addressed to address the challenges mentioned in the advisory. In the absence of regulation, there may be the development of different approaches by MNCs to govern this issue, which would make it more challenging to govern and regulate.[12]

Fair Dealing Principle to Govern TDM

In India, the fair dealing concept stems from Section 52 of the Copyright Act.[13] The Doctrine of fair dealing is required to achieve harmony or balance between the Author's competing monopolistic interests and the general society's creative objectives. What is the nature of TDM activity, and does TDM fall under the purview of copyright at all? Although, there are specific circumstances defined under Section 52 of the Act where the fair dealing principle will be applicable. But Indian courts have interpreted the fair dealing principle very liberally, and it is fair to say that, to some extent, it may cover the copyright issues associated with TDM as well.

The TDM activity may fall under Clause 1(i) of Section 52 of the Indian Copyright Act. It is defined as "a fair dealing with a literary, dramatic, musical or artistic work (not being a computer programme) for Private use including research". TDM done by a person for individual purposes may be protected under Clause (p) of the Article as well. Here, the question arises about the status of the TDM done for commercial purposes. The TDM becomes important in India as Copyright exists in non-registered work, too.[14] This may make it very challenging for the firms that use data for TDM to find out the status of the data and whether it is protected under copyright or not. If yes, finding the copyright holder may create another challenge for the AI developers. This may cause a delay in AI development and act as a barrier.

TDM through the Lenses of Section 52 and the Fair Dealing Principle

In India, the Court examined factors such as "amount and substantiality of the portion used" and "purpose and character of the use" to apply fair dealing under Section 52 of the Act:[15] Generally, the basic concept courts consider is whether the purpose served by the subsequent work and prior work is the same or substantially different. If it is different, there will be no fair dealing.[16] As Lord Denning said, "It is impossible to define what is 'fair dealing.' It must be a question of degree.[17] Because of lack of statutory definition, it is a matter for the courts to decide what fair dealing" is on a case-to-case basis. Even though there are specific circumstances mentioned under Section 52, there are still instances where the Indian courts have interpreted this Section so liberally that made the doctrine applicable in other cases as well.[18]

Interpretation of Section 52 of the Act

Various cases also suggest that the legislative intention behind Section 52 of the Act is to take care of the rights of the society at large in India. In the Wiley Eastern Ltd. Case, the Court stated that the

purpose of Section 52 of the act is to protect the rights under Article 19(1) of the Constitution of India. This Section aims to promote research, private study, review and criticism.[19] The 2012 amendment to Section 52 of the Copyright Act that extends the scope of this Section shows that the scope is very wide and can be widened according to societal needs. There are many instances where Indian courts have referred to US cases and considered the 4-step test.[20] Indian courts decided the fair dealing case according to the nature of the problem.[21] Although, unlike the US, Section 52 of the Act provides specific cases when fair dealing will be applicable. But the court has interpreted it very liberally. Since the fair dealing principle is not specific, it may be applicable w.r.t TDM. However, the TDM can't be governed only by the fair dealing principle, as Copyright is just one aspect of the issue. 'Fair dealing' is multidimensional, not uni-dimensional, and should be viewed from the point of view of all three stakeholders involved: the authors, the owner, and the public.[22] This liberal approach opens the scope for the TDM to be governed under Section 52 of the Act.

Compared to other countries, the rate of copyright infringement is very high in India. Among the 53 countries, India ranks 43 in the case of IP rights enforcement, including infringement, the civil and criminal legal procedures available to copyright holders, and the authority to carry out border controls and inspections by customs officials.[23] This indicates the availability of infringed copyrighted material. Additionally, in some cases, even it was held that if the user knows that the source is an infringed copy, still its use is not bad faith when the user acts rationally that his/her use is a "fair use". If the use of the infringed subject is used for a transformative purpose it would be considered as fair use.[24] It may create challenges as the developers may seek the benefit of this to bypass the copyright owner's permission. This may promote the infringement and unauthorized use and sale of the copyrighted subject matter in the country. Here, the important questions to ask are: What is the status of the use of exclusive economic rights of the copyright holder for non-commercial purposes? What is the stand on the temporary storing of the data? What works would fall under the fair dealing doctrine of Section 52 of the Act and what not?

Reproduction of the dramatic, literary, musical or artistic work is an infringement under Section 2(m) of the Act if the reproduction is made in contravention of the provisions of the Act.[25] Under Section 52, storing any work in any electronic medium, including the incidental storage of any computer program that is not an infringing copy under the Act, does not constitute infringement if the copyrighted material is used for personal use, including research. Therefore, in case of the use of such material for commercial purposes, it would not apply.

Additionally, the incidental or transient storage of work/performance for any technical electronic transmission/communication process to the public is not an infringement under Clauses (b) & (c) of the Act. So, temporarily storing the data to perform the action permitted under Section 52 of the Act is not an infringement. Therefore, the use of the copyrighted work for no commercial purpose that falls under Section 52 of the Act may not be an infringement.

The answer to the last question would depend on a case-to-case basis. In the case of *Authors Guide* v *Google*, the Court decided based on the 'need' for the infringement and 'literal necessary' to achieve the result and held that Google Books aimed to give significant information about the books to the buyers.[26] In *A.V.* v *iParadigms*, the Court has relied on the purpose of the keeping of copies of the copyrighted material. If the purpose is not mala fide and necessary to carry out the operation, it is not an infringement.[27] We may say that the courts may take similar approaches in case any dispute arises in the case of TDM.

TDM activity is a non-expressive use of the copyrighted work, so the protection should not be extended to the idea behind the protected work, as it is a settled principle in India that there are no copyrights associated with the expression or idea. In the case of *Kartar Singh Giani* the High Court held that "there are two points connected with the meaning of the expression "fair", in fair dealing. First is "an intention to compete and to derive profit from such competition" and second is "motive of the infringer".[28] None of these two elements are present in the TDM case. So, in similar cases, TDM may not cause copyright infringement if used without consent. The Court may take years to settle these issues, and the AI industry is in such a phase that needs such issues to be settled to make TDM feasible without any issues. Hence, all these factors require the legislature or executive intervention.

TDM in other Countries

In the EU, law that that talks about this issue is 'The European Directive on Copyright in the Digital

Single Market (2019/790)'. Article 3 of the directive talks about the mandatory exception under the EU copyright law that exempts the reproduction of subject matter protected under copyright. This Article also exempts extraction of subject matter protected under the sui generis database, SDGT. In the same way, Article 4 is available to every beneficiary. The beneficiaries can access any type of data except the cases where the Copyright owner has decided to opt out of this exception as Article 4 provides such privilege to the copyright holder.[29]

The US Copyright Office in its very recent 'Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence' provides that the person applying has to disclose the use of AI-generated content in the application with a brief explanation of the contributions of the person to the work. Since the exploitation under TDM is not an act of enjoyment, it can't be perceived as infringement. Even after such narrow exceptions, several issues may create hurdles for the EU exceptions to be implemented.[30]

In the EU, Article 3 exceptions are available only to research organizations and cultural heritage for scientific research. Article 4 also gives the right holder an option to opt out of the exception provided under this article. Because of this, the scholars argue that the exceptions are not wide enough to let R&D happen in the field of AI. Scholars have stated that the formulation of the copyright exception under these laws may be conceptually flawed theoretically even if it is underpinned by strategic policy goals.[31] The statistics show that the EU is lagging in the TDM patents as compared to the US and Asia (majorly China) and the major three factors are legal uncertainty w.r.t. to the treatment of activities under TDM in the EU and its Copyright law, lack of awareness and skill and infrastructural challenges.[32] This may keep the cost of AI development and can act as a hurdle for AI development.

The Japanese Government on 18 May 2018 amended the Japanese Copyright Act that came into force on 1 January 2019.[33] Compared to the EU, in Japan, the exceptions are framed liberally, and the Japanese government has ensured that copyright should not be an obstacle to the scientific development of AI. Unless it unreasonably jeopardizes the interests of the copyright holder, TDM is permitted without prior consent from the applicable rightsholders if the aim of the exploitation is not enjoyed by the person exploiting nor causing another person to enjoy the work unless the exploitation is causing harm to the interest of copyright holder."[34] Japan's TDM exceptions are very broad as they apply to both commercial and non-commercial purposes, and it apply without the consent of the right holders. In other words, exploitation is permitted in all the cases. In India, several socioeconomic factors encourage India to opt for liberal TDM exceptions, unlike the EU. The AI sector in the domestic economy in India is in a nascent stage and the competitive market demands the easy availability of data for AI development, otherwise, it would act as a barrier and cause delay in the development and due to the application of AI in other sectors, it may cause a spillover effect on the other relevant sectors as well. Additionally, linguistic, social and cultural diversity also may create serious challenges for society if data availability is restricted due to strict TDM exceptions.

Why liberal TDM is important for India?

AI is still in a nascent stage, and it is challenging to build clear long-term regulatory methods. Because of this, regulating AI is a difficult problem. Predicting how Copyright Law will respond to AI-generated work is not easy. It can be seen that the legal framework governing copyright laws will undergo significant transformation to cover and address these issues.[35] The EU prioritizes the protection of its citizens from harm in its draft of the AI Act.

Economic Factors

The ICRIER suggest that "the growth coefficient suggests that on average a unit increase in AI intensity, measured as the ratio of AI to total sales, can return USD 67.25 billion or 2.5% of GDP to the Indian economy in the immediate term. Further, this increase in investment will lead to an approximate 1.3 times increase in AI intensity, translating into spillover benefits of USD 85.77 billion for the Indian economy (3.2% of GDP)."[36] The 161st Parliamentary Report acknowledges the demonstrably significant contributions of artificial intelligence (AI) and its applications to the generation of revenue and the overall economic health of the Republic of India, coupled with its demonstrably positive impact on the nation's technological innovation capabilities, necessitate the controlled and secure expansion of AI within the nation.[37] This report also acknowledges that AI is capable of adding USD 957 billion by 2035 to the Indian economy.

Because of the availability of the data, big firms like Meta, OpenAI and Google, and already have a competitive advantage over the MSMEs in innovation and R&D in the field of artificial intelligence.[38] Easy excess of data to train the AI model would help these companies to keep the production cost low and this would put the MSMEs in a disadvantageous position.[39] Although there are significant opportunities for developing countries arising from AI-based technologies this new wave of technological change might bring significant challenges.[40]

The MNC would establish their base near to consumers and may result in these MNCs moving from developing to developed countries (for luxurious goods).[41] To keep the MNCs in the domestic economy the country needs to develop AI solutions in sectors other than manufacturing (services, shipping etc.) that would make it reasonable for these firms to keep their manufacturing base in India. On the other hand, services that are becoming digitally tradable and are highly labour-intensive especially the medium- to high-skilled labour services, AI and digital technologies, would help the MNCs to localize their firms from advanced economies to countries with cheap skilled labour.[42] As the services sector contributes about 50% of the Indian GDP, it is an important component of the Indian Economy.[43] In such a situation, it is important to develop infrastructure, governance and regulation to conceive AI adoption in the Indian service sector. This would need a low-cost feasible way to develop AI solutions for the services sector.

AI solution development needs to be developed in every sector that would compensate for the labour factor with other important factors like cheap manufacturing, cost-effective mobility/transport, and AI-enabled infrastructure that would keep the MNCs attracted to Indian domestic sectors despite the labour factor. To develop these solutions, AI firms need the policy space to develop their products. Such policy space can be created only if the data is easily available, the solution is accurate and unbiased, and the development cost is less to make these solutions competitive and affordable.

The nature of data matters a lot in TDM. More data means more precise recommendations by the AI output. However apart from the volume, data variance also matters for accurate results of an AI solution. This means that unrestricted access to more data would promote more competition and result in greater

economic gains for consumers of the country.[44] However, this needs to be taken care that the easy availability of data may also result in the greater misuse of data. All these insights give support to the liberal TDM exceptions for India.

Linguistic Factor

The AI language models are generally trained in the leading languages like English by the larger technology companies. The AI language models are still few and far behind the term of reference. This mandates a linguistically diverse country like India to progress in R&D in the AI sector to develop its own AI language models that are trained in the Indian languages. India is a very diverse country in terms of language. Investment in language resources and data storage entities is important as every state has their language, and training AI on data available on these languages is very important.[45] According to the Indian Government census of 2011, there are 121 languages spoken by 96.71 per cent of the country's population.[46] The languages are already acting as a barrier to accessing technology in the different regions of the country. For example, Amazon Polly supports none of the Indian languages, despite supporting many other languages. Microsoft Azure, which has 45 language support options includes only three Indian languages (Hindi, Tamil and Telugu). There are 48 socially recognized dialects of Hindi alone. This makes speech in India highly colloquial, which may decrease the accuracy of the output of AI speech models.[47] For developing and deploying voice-based AI systems, sufficient high-quality data is needed.[48] These factors may result in several challenges for an AI solution trained on data available in India.

The nature and the number of parameters used to train the language model would also affect the output of the AI model. In some specific sectors, India may need to set up parameters of AI training models according to the Indian social and cultural conditions to produce appropriate and less biased results.[49] Since every region of the country speaks a different language, the availability of data on any region-specific subject would be limited and can be found only from that region. In such a case, all the data should be easily availability to perform TDM for better working of the AI solution.

Discrimination, Misinformation and Biased Outputs

Generally, AI bias can enter by three ways in an AI system: data set bias, model design bias, and ML self-

training bias.[50] AI language models trained on low-quality data could harm the online "commons".[51] The DeepMind has noted six risk areas associated with the AI language models. Three of these areas are associated with the TDM.[52] The risk associated with areas such as discrimination, misinformation and human-computer interaction would depend on the data used in the TDM process.[53] Countries have also considered these areas in their policy space for AI governance.[54]

All the seven identified principles by NITI Aayog, are dependent on the TDM.[55] In India, to mitigate major risks associated with generative AI, it is important to use and deploy generative AI safely and fairly, which would take care of human rights. To address these risks, it is important to take care of the unintended consequences such as discrimination, misinformation, and malicious use and the government should support businesses to implement it.[56] This cannot be done without having reasonable freedom for the companies to carry out TDM activities to build AI solutions. The TDM law and regulation also have implications on human rights like discrimination, equality etc. It is important to know that when the AI system is used by an entity having constitutional or legal obligations should be unbiased.[57] Right to information under Article 19 of the Constitution in India would mandate correct information to be biased and untrue.[58] Rights against discrimination are enshrined under Articles 15 and 21 of the Constitution. As discussed above, the existence of Section 52 of the Copyright Act is itself based on Article 19 of the constitution, and the liberal TDM may go against the intent of the lawmaker. So, intellectual property rights cannot override the fundamental rights under the Constitution.

Employment Issues

To build an Encompassing Data Strategy there is a need to improve India's capacity to make data publically available and to encourage reliable, unbiased and inclusive data-sharing practices.[59] Researchers have argued that, in terms of employment, LLM may benefit lower-skilled workers more. As a larger population comprised of low-skilled people in India, it is important to frame the policies of generative governance accordingly to benefit the larger proportion of the society. It is important to note that with the opportunities for low-skilled labour, it also poses certain risks to the labour market that need to be addressed.[60] To reap the benefits out of it, regulation needs to be there to address the challenges associated with it.

Other Factors

According to the Niti Ayog publication, the lack of an enabling data ecosystem is one of the major challenges to setting up a robust AI ecosystem and its application. Low-intensive research and high resource costs are some other major factors across some major focus sectors. The areas that are focused are healthcare, agriculture, education, smart cities and infrastructure, and smart mobility and transportation.[61] Since the strict TDM rules would increase the cost of AI development and research, it would affect the overall AI development in India.[62]

During 2021-2023 Indian Generative AI startups have raised only USD 475 million as compared to more than USD 19 billion globally. This shows that for the proper development of AI India needs to frame the cost-effective policies for R&D in this field. Strict copyright policies would result in the increased cost of TDM practices.[63] Since TDM has a direct impact on R&D and innovation, the regulations need to be drafted very cautiously. Strict TDM regulations would take up the cost of AI development and it may put India in a disadvantageous position in terms of competition with the major AI players like China.

Furthermore, while the EU's approach to regulating the aforementioned aspect of AI is certainly pioneering, may not be the most suitable exemplar for adoption by Asian nations.[64] The requirement under Section 4 of the directive could pose a significant challenge to the continued development of AI innovations in the EU. The present regime may take the development cost of the AI sector up and this may have a detrimental effect on the innovation and R&D in this sector. The long-term financial potential of licensing copyright-protected information for training may be diminished as a result of this. Additionally, the procedure and the practical implementation of the reservation of rights remain ambiguous.[65]

However, there are several important provisions. In the EU that India may consider. For example, the EU AI Act requires AI companies to make publicly available a sufficiently detailed summary of the content used during the TDM process, and such provisions are not there in the abovementioned advisories of the Indian Government.[66]

Conclusion

As we know data or information is one of the most important factors for technological development and the purpose of IP should be to promote technological

development not to act as a barrier to its development.[67] The narrow approach of framing these exceptions is not appropriate and may have long-term implications. It may create a challenge for the development of the AI industry in the country. The AI industry is driven by Data which makes it very complex. The law should be made after a wide range of consultation from all the industries and a wide and long-term approach should be given to the policy. Too many regulations may cause challenges for the AI sector and affect the whole domestic economy.

India needs to draw a balance between the protection of the interests of right holders and promoting the R&D of AI. The biased AI solution created at the developer stage may not be easy to correct at a later stage. Since the data is not uniform, it is divided state-wide by language, peculiarities of the community, diverse religious beliefs and different casts, creeds and practices etc. It would make it challenging to collect such data for training AI solutions. If the data is less and not available, there would be more chances that the output of the AI system would be discriminatory, biased, and inaccurate. This would go not just against the fundamental rights of the Indian constitution but also international laws like principles of IP under TRIPS agreement Articles 7 and 8. If TDM is regulated strictly, it may go against the objective of Section 52 and the fair dealing principle.

The development of generative AI is anyway going to happen either with a liberal copyright exception with fewer risks and more opportunities and inclusive development or with strict copyright exceptions or comparatively more risks and challenges. The choice is ours.

The arguments raised in this article would guide policymakers to consider the challenges and Indian societal peculiarities to be considered while working in the policy space for AI governance in the IP field. This paper hopes that the Indian government and countries where similar situations of linguistic and social diversity are present to consider domestic peculiarities before relying on the practices and policies adopted by the countries.

References

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 March 2023).

3   Article 7 and Article 31 take care of the rights of the public at large and restrict the right of the IP holder in case of necessity.

4   Press Information Bureau, Existing IPR regime well-equipped to protect AI-generated works, no need to create a separate category of rights, https://pib.gov.in/PressReleasePage.aspx?PRID=2004715 (accessed on 10 March 2024).

5   The Copyright Act, 1957, Section 2 (d) (vi).

6   Advisory, Government of India, Ministry of Electronics and Information Technology, Cyber Law and Data Governance Group, No. eNo.2(4)/2023-CyberLaws–3 (accessed on 15 March 2024).

7   The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Section 3 (1) (b) (iv).

8   *The Chancellor, Master and Scholars of the University of Oxford & Ors.* v *Rameshwari Photocopying Services and Anr.,* Cs (Os) 2439/2012; *India TV Independent News Service Pvt. Ltd. & Ors.* v *Yash Raj Films Pvt. Ltd. & Anr. (Delhi High Court, 2* AIR 2013 (NOC) 315 (DEL.), *Devendra Kumar Ramchandra Dwivedi* v *State of Gujarat and Ors*. MANU/GJ/0440/2009.

9   The Ministry of Electronics and Information Technology, *Report of the Committee – D on Cyber Security, Safety, Legal and Ethical Issues*, https://www.meity.gov.in/artificial-intelligence-committees-reports,2019,28.

10   Information Technology Act, 2000.

11   Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011.

12   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 7 November 2023).

13   The Copyright Act, 1957, Section 52.

14   The Copyright Act, 1957, Section 45, states: ‘The authorization granted by the use of the word "may" implies the authors hold the unfettered discretion to pursue copyright registration’; In the case, *Sanjay Soya Private Limited* v *Narayani Trading Company,* 5011 OF 2020, the Bombay High Court has determined that the registration under the Copyright Act is not mandatory and is solely at the discretion of person. Furthermore, none of the provisions of the Copyright Act mandate the registration as a prerequisite for seeking legal remedies under the Act. Question 4 at https://copyright.gov.in/frmfaq.aspx.

15   *Civic Chandran* v *C Ammini Amma,* [1996] PTC 16 670.

16   *ESPN Stars Sports* v *Global Broadcast News Ltd and Ors*, 2008 (36) PTC 492 [Del] [Para 17].

17   *Hubbard* v *Vosper,* CA 1971 [1972] 2 WLR 389.

18   [India TV] *Independent News Services Pvt. Ltd* v *Yashraj Films Private Limited & Super Cassettes Ltd* FAO (OS) 583/2011; the Court gave relief to the appellant by applying fair dealing even though the subsequent use of the work does not come within the exhaustive list of cases provided under Section 52 of the Act.

19 *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".

20 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.

21 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.

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

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

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

25 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".

26 *Authors Guild* v *Google* 721 F. 3d 132 [2nd Cir. 2015].

27 *A.V. ex rel. Vanderhye* v *iParadigms,* L.L.C., 562 F.3d 630 [4th Cir. 2009].

28 *Kartar Singh Giani* v *Ladha Singh,* [1934] SCC Online Lah 277.

29 Directive of the European Parliament and of The Council of [2019] (EU) 2019/790.

30 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.

31 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.

32 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.

33 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.

34 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.

35 INDIAAI, Impact, opportunities and Challenges of generative AI, *A MEITY, NEGD & NASSCOM Initiative*, 38.

36 Kathuria R, Kedia M & Kapilavai S, Implications of AI on the Indian economy, *Indian Council for Research on International Economic Relations (ICRIER),* 2020.

37 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.

38 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.

39 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.

40 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.

41 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.

42 OECD *Initial Policy Considerations for Generative Artificial Intelligence*, OECD Artificial intelligence Paper series No-1, September 2023 at 21.

43 Ministry of Finance, Services contributed over 50% to GDP, 2022, https://pib.gov.in/PressReleasePage.aspx?PRID=1793804 (accessed on 15 November 2023).

44 Policy Department, Directorate-General for External Policies, Directorate General for External Policies of the Union PE 653.617 - November 2020, 7.

45 OECD, AI-language models: Technological, socio-economic and policy considerations, *OECD Digital Economy Paper*, April 2023, No. 352, 32 & 38.

46 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.

47 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.

48 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.

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

50 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.

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

52 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.

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

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

55 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.

56 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.

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

58 The Constitution of India, Article 15.

59 Kathuria R *et al*., Implications of AI on the Indian economy, *Indian Council for Research on International Economic Relations (ICRIER),* 2020; To build an Encompassing Data Strategy is one of the five key recommendations of ICRIER.

60 OECD, Initial Policy Considerations for Generative Artificial Intelligence*,OECD Artificial Intelligence Paper Series No-1,* September 2023 at 22.

61 NITI Aayog, National Strategy for Artificial Intelligence, June 2018, 24 & 46.

62 Margoni T, A deeper look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology, *GRUR International*, 71 (8) (2022) 687.

63 NASSCOM, *Large language Models (LLMs)- A Backgrounder'* Delloitee, 2023, 7.

64 Dermawan A & Mezei P, Artificial intelligence and consensus-based remuneration regime in Southeast Asia, *SSRN,* https://ssrn.com/abstract=4625850 (accessed on 7 November 2023).

65 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*, 19 (5) (2024) 453.

66 EU Artificial Intelligence Act, S. 1(d), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

67 Article 7 of TRIPS Agreement.