



Artificial Creations, Real Rights: Examining Copyrightability of Generative AI Data

Sunidhi Setia[†] and Madhumitha Vasan

Jindal Global Law School, O P Jindal Global University, Sonapat - 131 001, India

Received: 23rd December 2024; revised: 1st October 2025

Since the launch of ChatGPT by OpenAI, the investment in generative AI has quintupled. While individuals employing these tools are creating novel work every day, the parent companies are generating billions of revenues. With increasing number of high-profile lawsuits relating to Generative AI and Copyright around the world, the paper discusses the concept of copyrightability in the new era of technological advancements. Generative AI undergoes massive training using data to understand images, sounds and visuals. They then respond when prompted by a user.

Through this paper the authors discussed the basic concepts of ownership and then examine different scenarios to determine authorship over works generated by AI. Building on established concept like the essential need of human authorship and creativity element to grant copyright to the works of AI, the paper explores and divulge into new doctrines like the “Significant Input Test”. The authors argue that while copying large amount of copyrighted content off the internet and generating new content makes out a case for copyright infringement, the generative AI algorithm is premised on “transformative use” which in turn constitutes fair use. The authors did a comparative study of judicicia-legal development around the concept of copyrightability of the content produced by Generative AI focusing on jurisdictions like US and EU, suggesting a way forward for Indian legislature.

Keywords: Copyright, Generative AI, Authorship, Infringement, Creativity

With gigantic investments and ever-increasing copyright litigations, the question of exploring ethical and legal implications of copyrightability of AI generated contents becomes of paramount consideration. Considering the stakes of both innovators and creators the issue requires legal experts, policy makers, and leaders to join hands aiming balancing the rights of stakeholders. From copyright lawsuits files by the famous authors like John Grisham, George R.R. Martin to big media houses like ‘*The New York Times*’, the legal battles are increasing with advance tools and technology.

Content creators such as news providers, artists and authors across the globe have raised concerns that their work has been used in an unlawful manner to train AI models. An exception would be Open AI taking a license in July 2023 for a part of the text archive. However, this is nothing in comparison to the large volumes of texts, literary material and artistic works such as images and visual recordings taken off the internet without consent. The content used unlawfully to train AI models is infringing the copyright as well as the database rights of the content

creators. The recent case of *Getty Images Inc. v Stability AI Ltd.* is an important case study to understand the intersection of AI and related Intellectual property Issues. Here Stability AI had downloaded and stored large amount of copyrighted content of Getty Images on servers in the UK for training the AI model, Stability Diffusion. They had also infringed the right to communicate to the public by making Stability Diffusion widely available in the UK. Stability Diffusion uses the text and image prompts to produce synthetic images which is a replica of the copyrighted works hence constituting infringement.

With increasing number of high-profile lawsuits relating to Generative AI and Copyright around the world, the paper discusses the concept of copyrightability in the new era of technological advancements. Through this paper the authors discussed the basic concepts of ownership and then examine different scenarios to determine authorship over works generated by AI. Building on established concept like the essential need of human authorship and creativity element to grant copyright to the works of AI, the paper explores and divulge into new doctrines like the “Significant Input Test”. The paper

[†]Corresponding author: Email: sunidhisetia7@gmail.com

focuses on various issues related to intellectual property more specifically relating to copyright like ownership, licensing, infringement, and fair use to name a few. The paper will now discuss the foundational concepts relating to generative AI and the copyright law for authors to understand the subsequent arguments better.

Understanding Generative AI

“Generative AI” is a technology based on artificial intelligence, capable of producing text, videos, sounds and images. The data generated is of high quality and produced within a short time-frame. It was launched in 1960’s through the medium of chatbots. However, only in 2014, with the help of GAN’s (Generative Adversarial Networks), an algorithmic tool based on machine learning, generative AI became capable of creating images and videos seemingly authentic to the public. While this technology opened up a plethora of opportunities for easy learning and its usage in dubbings, it also unlocked gateways to deepfakes, cybercrimes, hacking and other nefarious activities.¹

The advancement of generative AI further led to generation of “transformers” and “language models”. “Transformers” enabled training of larger models by processing large amount of information contained in a book and provided insightful answers. These advanced models could draw connections between texts across pages of a book as opposed to just focusing on a single sentence. The “large language models” (LLM’s) were capable of actively engaging in writing text, painting realistic images and generating graphical videos. These tools additionally find use in coding, drug designing, transforming businesses, altering supply chains and product development.

Earlier, tools such as python or any other computer readable language was essential to write applications and generate an AI response. Nowadays commands fed in English language stimulates automatic processing. The results can also be styled according to one’s needs. Language processing is required to transform letters into words, speeches and essays. Images could be converted into videos. Neural networks combined with decoders and encoders generate facial expressions and a whole range of synthetic data.²

Dall-E and Chat GPT are built on Open AI’s GPT systems. While Gemini is built on Google AI. Let us briefly discuss the working of these models to understand the technological concepts better.

Dall-E is a multimodal application which can draw networks between vision, text and audio. Dall E-2 is

the improved model which can generate images or a continuous imagery in several formats in response to a prompt.

Chat GPT is a chatbot powered by AI. Open AI has created an interface where users can interact with the chatbots and the responses can be fine-tuned. ChatGPT is based on a mechanism where the chat history with the users have been incorporated into its results to create a simulation of a conversation. Microsoft made a public announcement of its investment in Open AI wherein it incorporated a version of the GPT into its search engine “Bing”.

Gemini is a product of Google. Google took advantage of the transformer AI technology for processing content in the form of text or proteins or a code. Google’s Gemini competed with other chatbots and its launch was rushed. Gemini is built on the lightweight version of its LLM’s. However, a rushed launch resulted in a crash in the stock price and committed an error of suggesting that Webb telescope was the first to detect a planet of the solar system. Google then designed the latest version of Gemini built on the advanced language models PaLM 2.

Understanding the technological and foundational concepts relating to the functioning of Generative AIs and the LLMs, the next section of the paper discusses the relevant concepts of copyright law.

Copyright and Content Generated by AI

Copyright law is aimed at granting protection to original works of authorship and offers the creators the exclusive right to utilize and circulate their creations. Section 13(1)(a) of the Indian Copyright Act states that copyrights exist in original literary, dramatic, artistic and musical works.³ “Author” has been defined under section 2(d) of the Copyright Act. As per Section 17 of the Copyright Act,⁴ author is the first owner of the copyright. Specifically in computer generated works, the person causing the work to be created is the author. As section 13 emphasize the copyright subsists in an original work thus “Originality” is a very significant concept in copyright.

Originality as recognized by the Copyright Act, eliminates the chances of copying or reproducing a work without permission. Originality is an acknowledgment of the effort of the artists and a standard set for works to merit copyright protection. The doctrines dealing with the standards of originality that have developed over the years include; sweat of the brow, modicum of creativity and skill and

judgment tests.⁵ The UK Courts followed the sweat of the brow doctrine which gave importance to effort and diligence.⁶ While the US Courts established that for works to merit copyright protection, a minimal level of creativity is a must. There has however been a global shift in granting copyright protection from factual content to a work meeting the standard of minimal level of creativity.⁷ The intellect of the author, individuality, creative spark and personal touch has been determinative in granting copyright. However, when content is being created by machines using AI technology, determination of authorship becomes challenging. Machines process vast volume of data and generate new data which is either similar to or inspired from original data. The algorithms use certain complex mathematical functions to identify patterns and generate new content. The algorithms possess no sense of creativity or mental frameworks but they can create content similar to human created content.⁸ Hence the dilemma of granting authorship.

The copyright with regard to generative AI can be classified into two categories i.e. input based and output based. These two types chiefly include; copyright over work used to train the AI systems and the copyright over works generated by AI. To create AI that responds to user prompts, massive set of training data is required. The LLM's use voluminous data publicly available and copyrighted, to train the AI systems. Using this data without permission could amount to infringement. In the NY Times lawsuit case, the articles from the newspaper were generated verbatim by the ChatGPT.⁹ AI companies such as Meta and Microsoft played defense and claimed that the usage of material fell under "fair use" and are backed by a wide number of precedents. Fair use is a legal right or defense available against infringement under the copyright law. In order to promote free flow of information, public interest and stimulate public discourse the fair use doctrine allows the reproduction of copyrighted work without permission in certain cases like news reporting, criticism, teaching, comments etc. The major arguments with regards to the copyrighted work used by Generative AI is dealt in detail in the latter part of the paper.

Because AI models cannot own a copyright, Stephen Thaler ended up losing the case to seek copyright protection over work autonomously created by an AI system he had developed. But some argue that generative AI is very similar to a camera. Photographers receive credit for the pictures they capture. Similarly using AI is also a collaborative

process where the human uses AI as a tool to display his creativity. The output is the creative work of a human which merits copyright protection. US Copyright Office holds a stance that humans do not exercise much creative control over the output and hence cannot claim copyright.¹⁰ In distinguishing between "works" and "ideas", some may argue that users only provide the idea. Absence of a human author, makes the work non-copyrightable.

Machine learning could be supervised or unsupervised. Human intervention in feeding inputs in machine learning during the creation of content could make the humans authors or co-authors of the content created. Few countries have recognized the legality of data created by machines. For example, the EU Directives have laid down laws to attribute credit for work generated using AI tools.

Who is the Author of the Content Created?

The first and foremost important with regards to the AI generated content is to decide the authorship of the work.¹¹ The content generated using AI tools may make the machines the creators. However, the data and mathematical patterns fed as algorithms into AI systems would belong to humans or the organizations involved in content creation. In cases where the AI systems are being trained using existing copyrighted material to produce new content, the ownership criteria could get complex. In situations discussed several legal questions arises like who exercised the creativity and judgment in creating the content? Who should be the author or the work? The authors argue and stress that these are questions of paramount importance wherein legal and technical experts have to collaborate before conferring copyright protection to AI generated content.

The courts around the world have derived several tests to determine the question of copyrightability of the AI generated content based on statutory laws and precedential jurisprudence developed around the law. AI has spread into several areas of technology and commerce. Ignoring work generated by AI would be disastrous for a country like India which considers itself to be leading in innovation. However, while giving importance to works created by AI, it is important to acknowledge the creative process and human involvement.¹² Therefore, it is important to strike a balance between safeguarding the human creativity and copyrightability of AI generated content. This led to the genesis of the "Significant Human Input Test".¹³ The "Significant Input" Test is two parts test

to determine authorship. The first criteria being whether there was human involvement. The second criteria entail the extent of human intervention in the process of creation. While determining the “extent”, it is important to establish that the finished product would be non-existent in absence of human creativity, intellect, labor, skills and judgement. For instance, writing legal commentaries can be challenging. Hence an AI tool can be utilized to summarize the judgments and extract vital points from existing research work, allowing the author to work efficiently. The author now wishes to improve the vocabulary and grammar of the paper using ChatGPT. The ChatGPT can only assist with structure and grammar and not improve the quality of the paper. The author however is entirely dependent on the AI assistance to enhance the quality of the paper. Can the author now claim copyright over this work? The answer is yes. This is because although ChatGPT was used, the author fulfills the twin requirement of the “Significant Input” test. There was human input in the creative process of shortlisting case laws and identifying relevant research work, Secondly, he used significant amount of labor and skill, to the extent that the work would not exist if not for his cognitive abilities to draft the paper.

The US utilizes a “Zarya” standard.¹⁴ This test was derived from the Zarya of dawn case. In this case the copyright was granted to a graphic novel although the imageries in the book were generated using AI. The copyright office based its decision on the skill and judgement exercised by the author in selecting and editing the images generated by AI.

While the “Significant Input” test checks for the extent of human involvement in the creation process, the “Zarya” standard is restricted to acts of “arrangement and selection” by the author in creating the graphic novel. A broader standard of creativity, skill and judgement is sufficient to evaluate the overall human involvement as per the “Significant Input” test. It does not pick out specific acts of arrangement and selection of pictures as measure of human skill. Therefore, due to the broadness in scope, the test of “skill and judgement” is better suited for the Indian Copyright law.¹⁵

To understand what qualifies as a “work of the author” or “creativity” some argue that content generated by AI lacks the human mind and creative powers resulting in loss of originality.¹⁶

Positioning across the Globe

Let us understand the global positioning with regard to copyrightability of works created by

Generative AI. Since most of the prominent cases in the area have been litigated in the US, we shall specifically draw our attention to the US. The US copyright office has issued statements that it would not offer copyright protection to works created by “non-human” authors, including content created by AI. In 2018, the Copyright office received an application based on visual work which the applicant described to be an autonomous work created by an algorithm which was running on a machine. The applicant was denied copyright protection because as per the representation in the application, no element of human authorship could be demonstrated. There were several appeals, after which the Review Board finally came to a conclusion on the matter that work is non-registerable as it was made without the “creative contribution of the human author”. More recently, in the year 2023, the Office received an application for work with human authored portions combined with AI generated images. It was graphic novel wherein the question was whether human authored text along with images generated by the AI Midjourney, could be copyrightable.¹⁷ The Office concluded that it is worthy of protection, however the individual images cannot be given copyright protection. There have been several applications ever since then including the works of AI in the work created by humans. The type and extent of the AI inputs haven’t been disclosed. However, it has been mentioned that it was co-authored by AI, or mentions have been made under the “Note to Copyright Office” section of the application wherein there have been indications to prove that work was completed with the assistance of AI. Certain other applications haven’t included the AI generated works, but have named the AI technologies used to complete the work in the “title” and “acknowledgment” sections of the application. Office has reached a standpoint where it has concluded that guidance of the public is essential in registration of works containing AI generated content. The office is seeking public guidance in issuing legal policies on how the law is to be applied in cases of utilization of copyrighted works in the training of AI tools and subsequent treatment meted out to the output.

The Office has made one thing clear that the term “author” as defined in the copyright law as well as the constitution excludes the contribution of “non-humans”. The Supreme Court has excluded the “non-humans” to interpret the Congress’s Constitutional power in granting the “authors” an exclusive right over their writings. In the case of *Burrow Giles Lithographic Co. v Sarony*¹⁸, the defendant was accused of making unauthorized copies

of certain photographs. To this, the defendant had replied that extension of the copyright protection to photographs by the Congress was unconstitutional, for the very reason that photographs are a product of the camera and not writings authored by a human. The Court held that Constitution's Copyright laws would extend to photographs because they are a representation of the intellectual conceptions of the author which are original in nature. The Court further defined "author" as someone to whom a work of science or literature owes its origin. He is an originator or a maker of a work. Authors are generally referred to the category of "persons" and the authorship is an exclusive right being granted to a human being to protect a work of his intellect or genius. Federal Appellate Courts have also arrived at the same conclusion as regards authorship. The Ninth Circuit has in fact held that a book containing words authored by a non-human spiritual being can be eligible for copyright protection only if there is involvement of a human selection and arrangement of revelations in the book. To emphasize that human authorship is essential to merit copyright protection, it is important to note that monkeys are excluded from being conferred a copyright for capturing photos with a camera. The Copyright Act very clearly refers to author's "children", "grandchildren", "widow" and "widower" which are all terms inclusive of humanity and excluding animals. The Compendium of Copyright Practices also stated that materials would not be eligible for copyright protection if it did not owe its origin to a human agent. The second edition of this Compendium also state that "authorship" meant for a work to be copyrightable, it must owe its origin to a human being. The present edition of the Compendium now states that for a work to qualify as a work of authorship, it must be the creation of a human being. Further, it explicitly states that works created by a machine through a mere mechanical process or in an autonomous or random fashion without the element of creativity or human input cannot be registered as a copyrighted work. The question essentially boils down to whether the work considered to be authored by a human was done with the assistance of computer tools or whether the traditional components of authorship such as the literary, artistic, musical expressions and the act of selection or rearrangement was done by a machine and not by a man. For all the works claiming copyright protection through human authorship, the

office will now consider whether the work created was through the contribution of AI in the form of mechanical reproduction of content or whether the human attributed his mental conception to the work and gave it a visible form. This is entirely based on the type of the AI tool used and the extent to which it contributed in the development of the work. A case by case analysis is required for the same. If the traditional components of authorship itself were a result of machine creation, copyright protection may not be conferred to it. For instance, if the AI technology receives a prompt from a user and generates all the literary, artistic, musical or other visual content as a result of the prompt, the traditional elements required for authorship were generated by a machine as opposed to a human. The Office is of an understanding that the AI technologies presently available do not give user the creative control over how the AI interprets and generates content based on the user prompts. The prompts are more like a manual of instructions to a commissioned artist. The machines decode the instructions but the manner in which it processes the output is left to itself. For instance, if the prompt by a user instructs the machine to write a poem about copyright law in the style of William Shakespeare,¹⁹ the machine will generate an output which identifies as a poem, making a mention of copyright and resembling the Shakespearean style. The technology would however be in control of the rhyme, words and structure and sequence of the text. When the AI is deeply involved in the creative structuring of the output, material generated is no longer the product of human authorship.²⁰ Therefore, such material is not worthy of copyright protection. However, if the skill of human selection and rearrangement is involved in the AI generated content, then the output is a result of human authorship and creativity.²¹ The artists may also modify content generated by AI in a sufficiently creative manner and introduce originality in it making it meet the standards of copyright protection. In the above cases, the copyright protection will only extend to the parts authored by humans or which are a result of human creativity. It will exclude all the AI generated content. The policy however does not entirely mean that tools of technology cannot play a role in human creativity. Authors have since a long time used such tools to create works to transform, adapt and recast their expressive authorship. A visual artist may use the technology of Adobe to photoshop

the image and still be the creative owner of the photograph. A musician may use guitar pedals to create recordings. What essentially is the differentiating factor is the extent of human creative involvement in the process and the output which formed the traditional building blocks of authorship. Copyright applicants however need to disclose the extent of AI involvement or assistance in the work created and also the human contribution in such works while registering their work for copyright protection. Such information is bearing upon the identification of the work or the existence, ownership and duration of protection of copyright.

Practical Application

In order to submit an application for works containing AI generated content, the applicant has to under “Author created” filed provide a brief about human contribution to the authorship of the work. They must use a Standard Application and clearly identify who is the author of the work. If AI generated content has been incorporated into a voluminous text in the work, the human must claim the human authored content in the textual work and differentiate it from the AI generated material.²² A clear distinction of authorship is essential. A human that creatively arranges the human and non-human content in the work should under the “Author created” field mention “Selection, arrangement and coordination of human authored content and describe the AI generated content separately”. The human cannot simply mention the name of the AI Technology or the name of the company providing the AI service and denote that the AI assisted in creation of content. A detailed demarcation of content is essential to decide on the authorship and extent of creativity and human intervention. If the AI generated content is more than de minimis, it should be excluded from the application itself. This can be done in the section on “Limitation of the claim” in the field of “Other” and under the heading “Material excluded”. Additional information may be provided in the “Note to CO” part of the application. Applicants unaware of how to fill the content, may simply mention that AI tools were used in generating the content, the copyright office will call the claimant to get clarity if necessary. In cases where no questions are raised on human authorship, despite the use of AI tools, nothing is required to be disclaimed in the application. If the work contained AI generated content, and if the

application pending before the office had does not disclose it, the copyright office can be contacted and informed about the same. The Examiner at the stage of review will coordinate with the applicant to understand the nature and extent of human authorship. For applications already registered, a supplementary registration can be done to correct any information in the application or enhance the information provided in the registration. In the supplementary registration, original material contributed should be in the “Author created” field, material generated by AI should be in the “Other/Material Excluded field” and the added material should be in the “Other/New Material Added” field. As long as there is human authorship, the office will publish the supplementary registration certificate with the disclaimer bearing AI generated content. Applicants failing to update the public records after obtaining registration for works including AI generated content may lose the registration. If the office becomes aware that the information about AI generated content has not been mentioned, it may take steps to cancel the registration. The Court may disregard the registration in cases of infringement especially if it is brought to the knowledge of the court that applicant knowingly withheld information or submitted inaccurate information, furthermore if the applicant knew that providing accurate information would have resulted in refusal of registration.²³

The EU Directives also includes clauses which suggest that copyright protection cannot extend to works created by “non-human” authors. The EU AI Act also mandates that data used by AI for training must be disclosed and cannot be used without authorization of the copyright owner. However, the extent to which copyright protection can be extended to works generated by AI still remains a question. While the EU Courts have held that copyright can only be conferred upon human authors, it has not eliminated the possibility of AI assisted works meriting originality if influenced by human creativity. In absence of an explicit statutory backing, companies must tread with caution while generating marketing content using AI as it may affect the brand. They must exercise caution while creating logos, slogans and images of the brand using AI as these are integral to boosting the brand value. Where there is absence of clear legislations, it is important to ensure compliance. Rouse advises companies to understand the application of copyright law during the creation of brand material.²⁴

Two specific issues that arise with regard to copyright and AI are as follows:

- (i) The chances of copyright infringement through usage of generative AI tools, wherein copyrighted material is used in the training of algorithms. This act of data collection is referred to as “mining”.
- (ii) Whether works produced through generative AI are copyright protected? and if yes, who owns the copyright?

With regard to the first aspect, the EU institutions have already taken a stance. Directive 2019/79/EU on copyright in the Digital Single Market, introduced the definition of “data mining” with a few exceptions. These exceptions apply while generative AI systems use data for training AI systems. It is important to note that specifically data mining is considered lawful in the region of EU.²⁵

- (i) Exceptions to rights conferred under Article 5(a) and 7(1) of the EU Directive 96/9/EC, Article 2 of Directive 2001/29/EC and Article 15(1) of the DSM, are reproductions carried out by research organizations and cultural heritage institutions considering they have legal access to such data and other subjects to be data mined (article 3 of DSM).
- (ii) Exceptions to rights under Article 5(a) and 7(1) of the EU Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of the Directive 2009/24/EC and Article 15(1) of the DSM include extractions and reproductions of lawfully accessible works carried out by an institution, as long as the data “mined” was done lawfully and the owner of the copyright has not raised any restrictions on the use of the work for the purpose of data mining as per Article 4 of the DSM.

Both these exceptions are subject to Article 5(5) of the 2001/29/EC Directive, states that the exceptions will only be applied in special cases which are not in conflict with the regular exploitation of the work and do not cause unreasonable prejudice to genuine interests of the right holders.

The second aspect dealing with ownership of works generated by AI is a little complex. In Europe, copyright protection only extends to works which are original and the intellectual creation of the author as opposed to a copy of the pre-existing work. There are limited exceptions to works owned by an author, such as works created by the employee which are held by the employer during the course of employment.

Copyright law is more of a national law and hence may differ in specifics among member countries. Hence the question remains as to who owns the copyright in AI generated works with human input, for which we must refer to the AI Act of the EU.

- Recital 105 of the AI Act,²⁶ refers to the relevance of data mining exceptions to copyright as per the Directive 2019/790.
- Recital 107,²⁷ requires that providers of the AI generative models need to provide a detailed summary of the content that is used for the purpose of training algorithms. This will allow parties with legitimate interests to pursue their rights under the European law.

Therefore, as we can see the AI Act does not handle specific issues of ownership in case of works produced using generative AI systems.

The Indian legal standing is quite complex. In 2020 an application attributing authorship to an AI “Raghav” was outrightly rejected whereas offering co-authorship to AI and humans was very well accepted.²⁸ In the said case, the court initially dismissed the argument that the AI (Raghav) was itself capable of creating something independently and hence could claim authorship. The court believed that the AI systems lacked the ability to think or be self-aware. However, the involvement of a human could be a game changer as the product of a human – AI collaboration could merit copyright provided the human exercises significant creative authority over the generated output. This judgment is synchronous with the Berne convention which also advocates human authorship. Therefore, completely autonomous AI generated works cannot be offered copyright protection.

This discussion questions whether AI creates something of its own or does it make mere modifications to what is already existing. The Copyright Act, 1957 requires a certain degree of creativity to grant authorship. It also explicitly states that the work must be authored by a person to claim copyright. Chatbots cannot be considered as authors. Section 17 of the Indian Copyright Act requires persons to be authors.²⁹ The nature of copyrights is mostly “human centric”.

Further if a developer of the AI claims copyright, over the work generated by AI, it would depend on the “Terms of Use”. The “Terms of Use” policy for most of the AI such as ChatGPT does not assert its claims over the outcomes. For example, Microsoft Paint cannot claim a copyright over a painting

created. ChatGPT being a service provider, cannot claim its right over the work of art. The product i.e. the painting is the skill of the human. Hence the right over the creativity vests with the human. The product developed using Open AI will also be a result of how a human mind used that particular service.⁸

Risks that Accompany Use of Generative AI to Develop New Content

- (i) Who is the owner of the copyrighted work? Software provider? Company licensing the software? Employees using the software?
- (ii) Companies entitled to commercially exploit AI generated content? Implementation or combining AI generated content into self-creations? Risks of infringement by third party?
- (iii) Ownership of the works created using generative AI may be impacted based on three criteria:³⁰
- (iv) Laws of a particular jurisdiction governing AI and creations of works
- (v) Extent of human input and role of AI in processing the output. Allowing the AI to completely produce the new work is vastly different from using AI as a tool to review the work or make minimum modifications or adaptations.
- (vi) IP provisions under the terms and conditions agreement with the service provider.
- (vii) Therefore, copyright protection over generative AI may belong to:
- (viii) Creators of the algorithm who may want to retain rights over works created by the algorithm.
- (ix) User of the AI systems.
- (x) Absolutely no one (The works created using AI tools cannot be owned by anyone and must be disclosed or put in the public domain as per the T&C of the AI systems used)

As per the present situation specified by the standard T&C of an average AI service provider, user is considered the owner of the work. The user is also then allowed to commercially exploit the work. However, the software provider may wish to use the works created by AI to train the AI or retain the ownership rights of the work produced by the algorithm or at least reduce its commercial exploitation. Hence it is important to check the terms and conditions of the generative AI systems.³¹

The major concern is the permission to commercially exploit the work generated by AI. The user is allowed to commercially exploit the work if the terms and conditions and the national law permit the same. There are however circumstances where commercial exploitation may lead to copyright infringement.

- (i) Output may completely or partially infringe preexisting works used to train the AI systems.
- (ii) The T&C specifies that one may not commercially exploit the works created using the AI systems.

Hence the T&C as well as the national laws ultimately predict the commercial exploitation of the work. Therefore, it is important to note that the contract with the service provider mentions that the AI tools have been trained with licensed content. The producer must also allow free commercial exploitation of the work and the output must not be reused to train the AI systems.³²

Regardless of the geography, some of the best practices that can be followed may include, the extent of usage of AI generated content should be made clear so that one can gauge the impact of AI on the final product, in case a third party raises a challenge, it is best to ensure that AI training systems and processes used for content creation are well documented and running the user agreements with AI service providers to understand how the agreement correlates with the ownership and commercial exploitation of the IP.

Understanding of Fair use and Derivative works

It seems like AI generated content is claiming copyright on existing works. Could there be an element of fair use involved? For instance, music generated by AI systems would contain few musical notes derived from existing songs. Would this be derivative work and infringe on existing works? The principles of fair use involve purpose of the use, nature of the work, the extent or volume of the work for which copyright protection is sought and the possible effect it may have on the commerce. These factors have to be reevaluated in light of AI generated content. It has been asserted by Open AI that AI generated work is of “transformative use”. This is typically when something is added to what is already existing to change its purpose or give it a different character as opposed to being a mere substitute. AI is transformative in the sense of creating something useful and hence falls under “fair use”. A relevant case would be *Authors Guild Inc. v Google Inc.*³³ Google created a digital database after scanning a number of physical books. The users can enter keywords to locate the books. Google contended that it would constitute fair use as only a small number of books were searchable by the users. 8000 lawsuits were filed by authors of the books claiming

infringement of copyright. The Court however held in favor of Google and stated that it was “transformative use” of the existing copyrighted texts. It enhanced the knowledge of searchers by providing information about the books without revealing too much about the copyrighted material.³⁴

Having said that, using enormous amounts of copyrighted data without prior permission or license, could lead to potential legal conflicts.

Infringement

With the increasing volume of data generated by AI systems, a vast amount of infringing data could be circulated globally. In the case of *TransUnion v Ramirez*,³⁵ a requirement was introduced by the Federal Court of US in order to establish a standing to sue for copyright. The plaintiff in the suit must demonstrate that they suffered “concrete harm” resulting from the infringement. Mere violation of the statute wasn’t sufficient to prove a case of infringement. This could change history in the field of copyright infringement for works related to AI. *TransUnion* could make it very difficult for copyright owners to bring claims against AI organizations for use of their copyrighted works as training data sets for AI without providing specific instances of infringement and the output as a result thereof. The copyright owners practically can satisfy the need of concrete harm by providing proofs of monetary harm, loss of financial opportunity, reputational harm. The owners can further prove concrete harm by producing evidence with regards to expenditure of time and resources

This judgment was influential in dismissing another case of *Raw Story Media v Open AI*. In this case the judge, Colleen McMahon had held that plaintiffs had failed to indicate the “concrete harm” caused to them as a result of Open AI removing copyright management information from their articles. Open AI has stated that these articles were used to train ChatGPT’s language model. Adopting the same reasoning for a wider number of cases could weaken the copyright infringement claims against use of copyrighted content for training AI models by AI companies. As regards the *Raw Story Media* case,³⁶ *Raw Story* and *AlterNet* had alleged that Open AI had violated the copyright law (DMCA) by using the copyrighted articles after removal of the names of the author and copyright notices in order to train their ChatGPT models. It was argued that section 1202(b) of the DMCA³⁷ was violated because the copyright

management information had been altered knowing very well that it would be instrumental in facilitating infringement in the future. Judge in the case had dismissed the allegation, because the plaintiffs had failed to prove a case of “concrete injury in fact”. This was vital to establish a standing under Article III as it sets the threshold. Plaintiffs could not prove that ChatGPT had disseminated their copyrighted works in response to queries by a user. The plaintiffs also failed to show significant harm or adverse after effects caused as a result of violation of the DMCA. The Judge, McMahon found that the claims were too abstract and did not meet the requirements for a federal standing. There was no scope for injunctive relief as ChatGPT models on a daily basis uses information scraped off the internet from innumerable sources for training purposes. In comparison to the voluminous content stored in the repository, the chances that ChatGPT would generate plagiarized content from one of the articles of the plaintiff is largely miniscule. As regards the violation of section 1202 (b) of the DMCA, in which the copyright management information is removed from copyrighted works and the copyrighted work is reproduced as is, the Judge was of the opinion that Generative AI models do not do that. The ChatGPT models apparently used the underlying information to synthesize new data and in turn generate novel responses to user prompts. The plaintiffs could not make a case of direct infringement or establish that the chances of infringement in the future were relatively high. In the case of *TransUnion v Ramirez*, the publisher’s claims were dismissed by Judge McMahon. A number of plaintiffs filed a case against the credit bureau called *TransUnion* as they were falsely labelled as terrorists and drug-traffickers in the credit reports. A right to sue was guaranteed to the plaintiffs under the Fair Credit Reporting Act by the Congress for such errors. However, the Supreme Court in this case established that this statutory right alone wasn’t sufficient to make a case at the Federal Court. As per the requirement under “Article III”, plaintiffs had to additionally prove “concrete harm” as a measurable injury. The Court only had time for actual controversies. Justice Kavanaugh held “No concrete harm, no standing”.³⁸ The impact of *TransUnion* could extend broadly even beyond the DMCA to acts of copyright infringement in the field of datasets used to train AI. Judge McMahon was of the clear view that without demonstration of harm by

showing that the AI generated identical or almost identical outputs, it would be difficult for copyright owners to battle out cases in the court when it comes to AI companies infringing copyrighted works. This judgment shakes the very foundation of claims being brought to courts by copyright owners. They have to completely rethink their claims when it comes to infringement caused by Generative AI. The Courts may not entertain cases without proof of “significant harm”. There would be no end to the amount of data the AI models could use for training and if legally permitted, the rights of IP holders could be violated widely without being able to seek legal recourse.³⁹ Such a precedent could be earth shattering for future legal disputes in the area of copyright infringement and permit the AI companies to closely mirror the expressive elements of the original copyrighted work to produce exactly similar “infringing” outputs.

Further, the copyright owners maybe incapable of detecting infringing data. The AI systems trying to mimic humans are able to generate accurate data and produce a replica of existing work. There is an urgent need for legal remedies. AI powered technological tools can also be availed for copyright detection.

Enforcement Mechanisms

AI technology produce vast amounts of infringing content using complex algorithmic systems. Automated take down systems removes potentially infringing content from the internet as soon as it detects it. These systems undergo training through AI powered machine learning and their scanning speed can make them detect infringing content faster than human reviewers. However, the drawbacks of relying on AI tools is that they could remove the legitimate content due to sudden inaccuracy in detection.

Several giants such as Microsoft and Open AI are facing lawsuits for copyright violation with regards to an AI tool Copilot which made use of licensed codes without acknowledging the owner. Stability AI and Midjourney are two AI art companies who have been sued for infringing the rights of artists by training their tools on images that were scraped.⁴⁰ Getty images also faced an infringement suit for training its AI art creating tools using millions of pictures without seeking permission from the owners of the artworks. Therefore, AI not only replicates copyrighted work, but its tools have also been trained on copyrighted works resulting in legal disputes.⁴¹

Conclusion

Based on the observations made throughout the paper, it appears that human authorship or significant human input and creativity is required for the work to be copyrightable. All said and done, purely machine generated work cannot be awarded copyright as a machine operating on an algorithm is not a legal entity or a natural person. Further, the way in which the AI tools process the user prompts are still not known to mankind. It is a complex set of neural networks mimicking the human brain. However no one has been able to decode how replicas of videos, images, recordings or visuals are produced by an AI tool. Further, one cannot be entirely dependent on an AI tool to produce authentic and reliable content. Although it works within a short time frame with high accuracy, it can never replace the human brain. It may process information at a faster rate, but may fail on the front of precision due to its limitations. As technology evolves, it is mandatory for law to parallelly evolve. That is exactly the reason why across the globe in different jurisdictions, legislations are being updated to factor in the technological advancements. While AI makes life easier, it poses several challenges. The US Copyright Office also requires a human to demarcate the AI input from the human input in a copyright application. Protection is ultimately awarded to the original expressions of ideas resulting from human creativity and intellect. A complete reproduction of preexisting content may not meet the standard of copyright protection. US requires a minimal level of creativity in the work for it to merit protection. EU directives also warrant the need for originality and work to be a reflection of the author’s intellect. Therefore, it becomes important to submit information regarding the amount of data on which the AI tools are trained to generate content. If minor changes are made by a human to content already generated by AI, there is no significant originality involved. Even going by the skill, labour and judgment test often relied on by UK courts to assess originality, if a machine generates most of the content without human effort and skill, it wouldn’t qualify for copyright protection. The standard for originality would be lowered if all AI generated content is given protection. We would also be facing a plethora of copyright infringement lawsuits if existing copyrighted works are reproduced without permission. The content seeking copyright protection must have some variations and alterations to make it transformative and distinguishable from already existing material. As in the Zarya of

Dawn case, “selection and rearrangement” done by a human may also be copyright protected while images generated by AI may not. Therefore, Indian copyright laws also need to be updated so as to avoid giving protection to purely AI generated content. The impact of AI tools on copyright laws could be far reaching and hence it is important to have sturdy copyright regimes. Authorship in case of AI generated content must be clearly defined in the legislation. The creators or owners of artwork must be duly compensated and the consumers must be made aware of what they must steer clear off. Litigation could be reduced. This works in the interests of consumers, creators and stakeholders.

References

- 1 IP and frontier technologies, Generative AI: Navigating intellectual property, (*WIPO*) <https://www.wipo.int/export/sites/www/about-ip/en/frontier_technologies/pdf/generative-ai_factsheet.pdf> (accessed on 23 August 2024).
- 2 Lawton G, What is generative AI? Everything you need to know, (*Tech Target Enterprise AI*) <<https://www.techtarget.com/searchenterpriseai/definition/generativeAI#:~:text=Generative%20AI%20is%20a%20type,imagery%2C%20audio%20and%20synthetic%20data.>> (accessed on 23 August 2024).
- 3 The Indian Copyright Act, 1957, s 13(1)(a).
- 4 The Indian Copyright Act, 1957, s17.
- 5 Kumar A, ‘Overview of tests of copyright protection in India’ (*iPleaders*, 17 May 2023) <<https://blog.ipleaders.in/overview-of-tests-of-copyright-protection-in-india/>> (accessed on 15 November 2024).
- 6 Nooniam M, UK’s sweat of the brow doctrine: Doctrine of originality in copyright, (*IIPRD Blog*, 16 April 2019) <<https://iiprd.wordpress.com/tag/uks-sweat-of-the-brow-doctrine/>> (accessed 15 November 2024).
- 7 Saraswat T, Evolution of tests of creativity in copyright, (*IPLF*, 3 May 2023) <https://www.ipandlegalfilings.com/evolution-of-tests-of-creativity-in-copyrights/> (accessed 15 November 2024).
- 8 Vishnu S, Navigating the grey area: Copyright Implications of AI generated content, *Journal of Intellectual Property Rights*, 29 (2024)103.
- 9 Audrey Pope HLR, *NYT v OpenAI: The Times’s About-Face* (*Harvard Law Review*, 10 April 2024) <<https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/>> (accessed on 15 November 2024).
- 10 Yoon J & Borges C, Informing the innovation policy debate: Key concepts in copyright laws for generative AI, (*CSIS*, 12 April 2024) <<https://www.csis.org/blogs/perspectives-innovation/informing-innovation-policy-debate-key-concepts-copyright-laws>> (accessed on 23 August 2024).
- 11 Rana L, Artificial intelligence and copyright – The authorship, (*Mondaq*, 18 December 2019) <<https://www.mondaq.com/india/copyright/876800/artificial-intelligence-and-copyright-the-authorship>> (accessed on 15 November 2024).
- 12 Manish Jindal ‘Artificial Intelligence and Copyright in India’ (*Bytescare Blogs*, 21 June 2024) <<https://bytescare.com/blog/artificial-intelligence-and-copyright-in-india>> (accessed on 15 November 2024).
- 13 Miles S, Copyright in works generated by artificial intelligence – Is human input necessary?, (*Edwin Coe LLP*, 8 November 2023) <<https://www.edwincoe.com/blogs/main/copyright-in-works-generated-by-artificial-intelligence-is-human-input-necessary/>> (accessed on 15 November 2024).
- 14 Mathur A, Copyright Registration for AI Generated Works: Zarya of the dawn and the dawn of a new creator, (*Center for art law*, 6 July 2023) <<https://itsartlaw.org/2023/07/06/copyright-registration-for-ai-generated-works-zarya-of-the-dawn-and-the-dawn-of-a-new-creator/>> (accessed on 15 November 2024).
- 15 Chhabra H & Pandey K G, Balancing Indian Copyright Law with AI generated content: The “Significant Human Input” approach, (*IJLT*, 26 February 2024) <<https://www.ijlt.in/post/balancing-indian-copyright-law-with-ai-generated-content-the-significant-human-input-approach>> (accessed on 23 August 2024).
- 16 Tyto, Generative AI and Human Originality, (8 February 2023) <<https://tytopr.com/blog/generative-ai-human-originality#:~:text=While%20AI%20generated%20content%20may,emotions%2C%20experiences%2C%20and%20imagination.>> (accessed on 1 November 2024).
- 17 Hookway J, AI-Generated Comic Book ‘Zarya of Dawn’ keeps copyright but key images excluded: Protection can’t be granted to images created with artificial intelligence software, U.S. Copyright Office says, (24 February 2023) <<https://www.wsj.com/articles/ai-generated-comic-book-zarya-of-the-dawn-keeps-copyright-but-key-images-excluded-c8094509>> (accessed on 1 November 2024).
- 18 *Burrow Giles Lithographic Co. v Sarony* [1884] 111 U.S. 53.
- 19 O&A, Copyright and artificial intelligence: Navigating a new frontier, (12 February 2024). <<https://oandapc.com/copyright-and-artificial-intelligence-navigating-a-new-frontier/#:~:text=Example%20of%20how%20copyright%20does,the%20structure%20of%20the%20text.>> (accessed on 15 November 2024).
- 20 Liu A H, Can A Machine Be Creative?, (*The Artificial Intelligence Blog*, 17 April 2023) <<https://blogs.duanemorris.com/artificialintelligence/2023/04/17/can-a-machine-be-creative/>> (accessed on 15 November 2024).
- 21 Lee M S, Artificial intelligence and copyright law: Less than meets the eye, (*Insights*, 29 August 2023) <<https://www.rimonlaw.com/artificial-intelligence-and-copyright-law-less-than-meets-the-eye/>> (accessed 15 November 2024).
- 22 Federal register : The daily journal of the United States Government, copyright registration guidance: Works containing material generated by artificial intelligence, (*Copyright Office, Library of Congress*, 16 March 2023) <<https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence#:~:text=When%20an%20AI%20technology%20determines,disclaimed%20in%20a%20registration%20application.>> (accessed on 15 November 2024).
- 23 Perlmutter S, Copyright registration guidance: Works containing material generated by artificial intelligence, (*Federal Register: The Daily Journal of the United States Government*, 16 March 2023)

- <<https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence>> (accessed on 28 September 2024).
- 24 White H *et al*, AI and Copyright: What materials can be used and what can be protected?, (*Rouse Insights*, 28 August 2024) <<https://rouse.com/insights/news/2024/ai-and-copyright-what-materials-can-be-used-and-what-can-be-protected#>> (accessed on 28 September 2024).
- 25 Directive (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA Relevance) (*Official Journal of the European Union*) <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0790>> (accessed on 15 November 2024).
- 26 The EU Artificial Intelligence Act, Recital 105.
- 27 The EU Artificial Intelligence Act, Recital 107.
- 28 Sarkar S, Exclusive: India recognizes AI as co-author of copyrighted artwork, (*Managing IP*, 5 August 2021) <<https://www.managingip.com/article/2a5czmpwixyj23wyqct1c/exclusive-india-recognises-ai-as-co-author-of-copyrighted-artwork>> (accessed on 15 November 2024).
- 29 The Indian Copyright Act, 1957, s17.
- 30 Appel G, Generative AI has an intellectual Property Problem, (*Harvard Business Review*, 7 April 2023) <<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>> (accessed on 15 November 2024).
- 31 Zirpoli C T, Generative artificial intelligence and copyright law, (*Congressional Research Service*, 29 September 2023) <<https://crsreports.congress.gov/product/pdf/LSB/LSB10922>> (accessed on 15 November 2024).
- 32 European Innovation Council and SMEs Executive Agency, ‘Artificial Intelligence and Copyright: use of generative AI tools to develop new content’ (*European Commission*, 16 July 2024) <https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/artificial-intelligence-and-copyright-use-generative-ai-tools-develop-new-content-2024-07-16-0_en> (accessed on 28 September 2024).
- 33 *Authors Guild Inc. v Google Inc.* (2d Cir. 2015) 804 F.3d 202.
- 34 Julia Yoon & Chris Borges, Informing the innovation policy debate: Key concepts in copyright laws for generative AI, (*CSIS*, 12 April 2024) <<https://www.csis.org/blogs/perspectives-innovation/informing-innovation-policy-debate-key-concepts-copyright-laws>> (accessed on 23 August 2024).
- 35 *TransUnion LLC v Ramirez* 594 U.S.(2021).
- 36 *Raw Story Media, Inc v Open AI Inc*, Civil Action 1:24-cv-1514-CM-OTW (S.D.N.Y. Apr. 10, 2024).
- 37 Digital Millennium Copyright Act, s1202(b).
- 38 Moss A, Will AI copyright claims keep standing after new ruling?, (*Copyright Lately*, 10 November 2024) <<https://copyrightlately.com/raw-story-copyright-lawsuit-standing/>> (accessed on 15 November 2024).
- 39 Day J, Generative AI generates excitement – and copyright concerns, (*Insights*, 28 April 2023) <<https://www.jonesday.com/en/insights/2023/04/generative-ai-generates-copyright-concerns>> (accessed on 15 November 2024).
- 40 Dilmegani C, Generative AI copyright concerns & 3 best practices in 2024, (*AI Multiple Research*, 4 June 2024) <<https://research.aimultiple.com/generative-ai-copyright/>> (accessed on 23 August 2024).
- 41 Appel G, Neelbauer J & Schweidel D A, Generative AI has an intellectual property problem, (*Harvard Business Review*, 7 April 2023) <<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>> (accessed on 23 August 2024).