

THE COPYRIGHT PROBLEM WITH EMERGING GENERATIVE AI

APOORVA VERMA*

ABSTRACT

Generative AI, a rapidly advancing field in artificial intelligence, has gained significant attention and recognition for its ability to create original and diverse content such as images, music, text, and even entire virtual worlds. Intelligence (AI) technologies, such as ChatGPT, have ushered in a new era of content creation, enabling users to produce vast amounts of original text at an unprecedented speed. By leveraging complex algorithms and deep learning techniques, generative AI systems can autonomously produce new and innovative outputs that mimic human-like creativity. However, as this technology continues to evolve and become more pervasive, it raises various legal, ethical, and societal concerns. This technological advancement has brought forth a range of complex copyright issues that demand careful consideration. This paper examines the challenges and implications surrounding copyright in the context of generative AI. The study delves into the unique characteristics of generative AI, where the AI system autonomously generates creative content, blurring the lines of authorship and ownership. The concept of “originality” in the context of AI-generated works is critically examined, with a focus on the extent of human intervention and the transformative nature of the output. Furthermore, the paper analyzes the challenges posed by the lack of clarity regarding authorship and ownership, the enforceability of copyright laws, and the potential infringement risks associated with AI-generated content in the context of India.

Additionally, the study sheds light on the responsibility of AI developers and platform providers, emphasizing the need for proactive measures to protect original creators and users from copyright infringement. It explores potential solutions, such as licensing mechanisms, attribution protocols, and technological safeguards, to strike a balance between promoting innovation and safeguarding copyright interests.

I. INTRODUCTION

The term “generative AI” refers to a category of artificial intelligence [“AI”] models and techniques that are designed to generate new and original content. Generative AI has created a lot of buzz around the world due to its ability to create new content such as images, text, music, or even videos, that resemble the patterns and characteristics found in the training data they were exposed

* 3rd Year Student at Gujarat National Law University.

with a substantial amount of data, which serves as the foundation for the model's learning.¹ This data can encompass various forms such as text, code, graphics, or any other relevant content related to the specific task at hand. In machine learning, three fundamental elements exist: models, data, and compute. Models are algorithms that take inputs and produce outputs. Sufficient and diverse data is crucial for the algorithms to generate valuable output.² Models need to be flexible enough to capture the complexity present in the data. Additionally, adequate computing power is necessary to execute the algorithms effectively.³

Once the training data is gathered, the AI model analyses the data patterns and relationships to comprehend the underlying rules governing the content. Through continuous learning, the AI model refines its parameters, enhancing its ability to mimic human-generated content.⁴ As the AI model generates more content, its outputs become increasingly sophisticated and persuasive.

Initially, a human input is required to provide a prompt to a generative model for content creation. Typically, creative prompts tend to produce creative outputs. As a result, the role of a "prompt engineer" is expected to emerge as a recognized profession, at least until more advanced AI models come into play. The field has already witnessed the development of an 82-page book dedicated to DALL-E 2 image prompts and the establishment of a prompt marketplace where users can purchase prompts from others for a nominal fee. Users of these systems usually need to experiment with multiple prompts before achieving the desired outcome. Generative AI models exhibit a wide range of capabilities and can process various types of content, including images, lengthy texts, emails, social media posts, voice recordings, program codes, and structured data.⁵ They can generate new content, provide translations, answer questions, perform sentiment analysis, generating summaries, and even create videos.

Training Models Of AI

The models can be mainly categorized into three types:

¹James Vincent, *The scary truth about AI copyright is nobody knows what will happen next*, THE VERGE (Nov. 15, 2022), <https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data>.

²What is generative AI, MCKINSEY & COMPANY, (Jan. 19, 2023), <https://www.mckinsey.com/industries/technology-digital-transformation/ai/generative-ai> | McKinsey.

³Stephen Amell, *How to train a Generative Model*, MEDIUM, (June 16, 2023).

⁴Bernard Marr, *The Difference Between Generative AI And Traditional AI*, FORBES, (July 24, 2023), <https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/>.

⁵Thomas H. Davenport and Nitin Mittal, *How Generative AI is changing Creative Work*, HARV. BUS. REV., (Nov. 14, 2022), <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>.

Contrary to the notion of conjuring material out of thin air, generative AI platforms undergo a training process using extensive datasets comprising vast amounts of images and text. These datasets are processed by software, constructing billions of parameters that enable the AI platforms to identify patterns, establish relationships, formulate rules, and make judgments and predictions when generating responses to prompts.⁹ Thus, the AI platforms leverage their learned knowledge and patterns to generate text outputs.

Generative AI models have been used in various applications, including:

1. **Image Generation:** Models like Generative Adversarial Networks (GANs) can generate realistic images from scratch, based on patterns and styles learned from training data.
2. **Text Generation:** Language models, such as GPT (Generative Pre-trained Transformer), can generate coherent and contextually relevant text, mimicking human-like language patterns and styles.
3. **Music Composition:** Generative AI models can generate original music pieces, imitating different genres or even composing unique melodies based on patterns learned from training data.
4. **Video Synthesis:** Advanced generative AI models can generate video sequences, altering or combining existing video content to create new, realistic videos.

Generative AI holds great potential in creative fields, content generation, and simulation scenarios. However, it also raises ethical concerns related to the potential misuse or creation of misleading or harmful content. It is important to consider responsible use and regulation when deploying generative AI technologies. The process of utilizing generative AI entails certain legal risks, particularly concerning potential infringement of intellectual property rights. These risks give rise to unresolved legal questions. For instance, there is uncertainty regarding whether copyright, patent, and trademark infringement apply to creations generated by AI. Additionally, it also raises the issue of determining the ownership of the content produced by generative.

IV. IMPORTANCE OF DATA OWNERSHIP

Ownership is a fundamental concept deeply ingrained in our daily lives and in the fundamental workings of society. It represents the allocation of rights and responsibilities for a particular property to an individual or organization. When we speak of property rights, we refer to the rights

⁹*Supra note 3.*

Copyright infringement is a growing concern within the AI community, leading governments to introduce regulations to address this issue and hold AI developers and their systems accountable. OpenAI CEO Sam Altman acknowledged the importance of government regulation during his testimony before Congress. He emphasized the need for responsible AI systems that respect the rights of original content creators and ensure proper attribution.¹² Generative AI technology has the capability to generate novel and ground-breaking content that goes beyond the limits of traditional methods. However, the application of this technology has raised significant inquiries regarding the copyrightability of the resulting creations. The complexities revolve around determining if the Intellectual Property (IP) generated through generative AI is eligible for protection under the current intellectual property laws and establishing ownership rights over such IP. Copyright law generally grants exclusive rights to the creators of original works, allowing them to control how their creations are used and distributed. However, when it comes to generative AI, questions arise about the authorship and ownership of the content generated by the AI system. In the case of AI-generated content, determining the original creator becomes complex. Traditional notions of authorship, where a human creates something based on their creativity and effort, do not easily apply to generative AI. Instead, the AI system generates content based on patterns and data on which it has been trained, without direct human input for each output.

This raises questions about who should be considered the author or owner of the AI-generated content. This question needs to be understood from two dimensions:

The concept of input and output copyright refers to the copyright protection of both the input data used to train an AI system and the output generated by the AI system.¹³

A. Input Copyright: Does The AI Training Process Infringe Copyright of Other Works?

1. Understanding Input Data

Input data refers to the information used to train an AI system. It can include text, images, videos, or any other form of content. Input copyright deals with ownership and rights associated with the data used to train the AI system. If the input data is subject to copyright protection, the rights and restrictions associated with that data would typically apply. The use of copyrighted input data

¹² Johana Bhuiyan, *OpenAI CEO calls for laws to mitigate 'risks of increasingly powerful' AI*, THE GUARDIAN, (May 16, 2023), OpenAI CEO calls for laws to mitigate 'risks of increasingly powerful' AI | ChatGPT | The Guardian.

¹³ Jessica Fjeld & Mason Kortz, *A Legal Anatomy of AI-generated Art: Part I*, (Nov. 21, 2021), HARV. J.L. & TECH, available at <https://jolt.law.harvard.edu/digest/a-legal-anatomy-of-ai-generated-art-part-i>.

copyright issues. Universal Music Group claims that AI companies violate copyright by using artists' songs in their training data.¹⁷

These arguments are currently being tested in court through multiple lawsuits. There are cases where artists have filed a class-action lawsuit alleging copyright infringement in the training of AI image programs. For instance, Getty Images¹⁸ has also filed a lawsuit against Stability AI, accusing the company of copying millions of copyrighted images from their websites for training purposes. These instances highlight how the lawsuits have been brought in the courts challenging the principle of fair use principle in the light of increased use of copyrighted technology by the AI. In a much recent case of *Andersen v. Stability AI et al.*¹⁹, filed in late 2022, a group of three artists formed a class to sue several generative AI platforms. The artists' arguments are based on the claim that these AI platforms used their original works without obtaining the necessary licenses to train their AI systems in a way that allows users to generate works that closely resemble their existing protected works. According to the artists, these generated works may not be transformative enough and could be considered unauthorized derivative works. If the court determines that the AI's works are indeed unauthorized and derivative, significant penalties for copyright infringement can be imposed.

B. Output Copyright: Does AI qualify as copyright owner for the contents generated?

1. Understanding Output Data

Output copyright pertains to the content generated by an AI system. This includes any text, images, music, or other creative works produced by the AI. The question of who holds the copyright to AI-generated output is more complex. As mentioned earlier, traditional notions of authorship may not directly apply to content generated by AI.

Some argue that if AI systems autonomously create content without human intervention, the output should not be eligible for copyright protection²⁰. Others propose that the organization or individual who owns and operates the AI system should be considered the copyright holder of the generated output. For example, OpenAI according to its Terms of Use appears to assign any copyright (rights, title and interest) to the user.²¹ While addressing the issue of assigning copyright

¹⁷Mia Sato, *Drake's AI clone is here- and Drake might not be able to stop him*, THE VERGE, (May 1, 2023), <https://www.theverge.com/2023/5/1/2>.

¹⁸ Getty Images (US) Inc. v. Stability AI Inc., No. 23-cv-135(D.Del.).

¹⁹ Andersen et al v. Stability AI Ltd, Midjourney, Deviant Art, 23-cv-201, (N.D. Cal.).

²⁰Michael Kasdan and Brian Pattengale, *A look at Future AI Questions For The US Copyright Office*, LAW360 LEXISNEXIS, (12 Feb. 2022), [Law360-A-Look-At-Future-AI-Questions-For-The-US-Copyright-Office.pdf](https://www.law360.com/legal/article/Law360-A-Look-At-Future-AI-Questions-For-The-US-Copyright-Office.pdf) (g2bswiggins.wpenginepowered.com).

²¹ See also OpenAI, Terms of Use, OPENAI (Dec. 8, 2020), <https://openai.com/policies/terms-of-use>.

works require a human author, works created by humans using generative AI could potentially be eligible for copyright protection, depending on the extent of human involvement in the creative process.

The recent developments in copyright proceedings and the subsequent Copyright Registration Guidance in US suggest that the Copyright is unlikely to recognize human authorship when an AI program generates works based on simple text prompts.²⁵

3. *The Case of Zarya The Dawn*

In September 2022, Kris Kashtanova registered a copyright for a graphic novel *Zarya the Dawn* that was illustrated using images generated by the AI system Midjourney in response to textual inputs. However, in October, the Copyright Office initiated cancellation proceedings after discovering that Kashtanova had not disclosed the use of AI.²⁶ Kashtanova argued that they authored the images through a creative and iterative process, distinguishing it from the case of Dr. Thaler, who attempted to register an image created by AI. Despite the argument, on February 21, 2023, the Copyright Office determined that the images were not eligible for copyright protection, as the images were produced by Midjourney, rather than Kashtanova whose role was limited to providing prompts and arranging the images, was not the mastermind behind the comic.²⁷

Based on this decision, the Copyright Office released guidance in March stating that when AI "determines the expressive elements of its output, the generated material is not the product of human authorship" and therefore not eligible for copyright protection.²⁸

In the case where AI-created works are deemed eligible for copyright protection, the question of ownership arises. The ownership is usually granted to the "author or authors" of the work. However, there have been no clear rules established by courts or under the Copyright regime regarding the identification of the "author or authors" of AI-created works.²⁹ However, the question that arises is: who is the author or creator of the output the user, the AI, the developer

²⁵ Joao Quintais, *Generative AI, Copyright and the AI Act - Kluwer Copyright Blog*, (May 9, 2023), KLUWER COPYRIGHT BLOG, <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>.

²⁶ Riddhi Shetty, *AI Comic Art Dispute Leaves Copyright Protections Open-ended*, (Feb.24, 2023), BLOOMBERG LAW NEWS, <https://news.bloomberglaw.com/ip-law/ai-comic-art-dispute-leaves-copyright-protections-open-end>.

²⁷ Blake Brittain, *AI – created images lose U.S. copyrights in test for new technology*, REUTERS, (Feb. 23, 2023), AI-created images lose U.S. copyrights in test for new technology | Reuters

²⁸ United States Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, VOL.88, March 16 2023, https://copyright.gov/ai/ai_policy_guidance.pdf

²⁹Simon Chesterman, *Artificial Intelligence And The Limits Of Legal Personality*, 69 INT'L & COMPARATIVE L. QUARTERLY, 819–844 (2020).

prepared a draft AI Act, and the United States has introduced an ‘AI Bill of Rights’³¹. Each jurisdiction's regulatory developments have been shaped by these unique contexts. Governments worldwide are actively engaging in regulatory efforts to address the challenges posed by generative AI, considering the purpose of regulation, liability frameworks, and important regulatory elements within their specific contexts but it has proved to be a very complex work that may need changing the way the copyright has been understood and perceived for so long.

A. Requirements Under Indian Copyright Act

In India the copyright is protected under the Indian Copyright Act 1957. Section 13³² of the Copyright Act of 1957 grants copyright protection for various forms of creative works, which includes literary works, musical works, theatrical works, artistic works, sound recordings, and cinematographic films. These categories encompass a broad range of creative expressions. For instance, the Act protects literary works such as books, manuscripts, and poetry. It ensures that original works in literature, drama, music, art, as well as cinematographic films and sound recordings, are safeguarded against unauthorized use and access. The Copyright Act of 1957 establishes provisions to prevent infringement and unauthorized exploitation of these protected works.

There are several issues within the current copyright laws that one is faced with while understanding the scenario created by generative AI. One key issue is that the Copyright Act of 1957 protects "original" literary and artistic works. However, according to a prevailing theory, AI is currently unable to create content that is truly “original.” Instead, AI-generated works are often seen as adaptations or modifications of existing information that the AI has accessed, analysed, and been trained on. This perspective stems from the understanding that AI systems rely on datasets that are influenced by the biases and limitations of their human creators. In other words, the output of AI is considered to be derived from pre-existing information rather than being genuinely novel or original.

The Copyright Act in India requires that a work must meet the test of a ‘modicum of creativity’ as established by the Supreme Court in the *Eastern Book Co vs. D.B. Modak*³³ case. The court held

³¹ Johana Bhuiyan and Nick Robins-Early, *The EU is leading the way on AI laws. The US is still playing catch-up*, THE GUARDIAN, (June 14 2023), <https://www.theguardian.com/technology/2023/jun/13/artificial-intelligence-us-regulation>.

³² The Copyright Amendment Act, No. 65 of 1984, § 13, Act of Parliament, 1984 (India).

³³ *Eastern Book Company and Ors. v. D.B. Modak and Anr.*, (2008) 1 SCC 1.

seeking clarification on the legal status of the AI named *Raghav*. This action suggested that the copyright was granted by error without application of mind moreover in the notice of withdrawal the onus was placed on the applicant to “inform the copyright office on the legal status of the AI Tool i.e., Raghav Artificial Painting App”.³⁹

AI is not currently recognized as a legal entity in India by any statute. As a result, the existing legal framework may not adequately address works where the actual creator is neither a human nor a recognized legal. Recognizing AI as an entity separate from a person and granting it ownership of intellectual property rights may give rise to potential copyright violations. Moreover, such infringements may not be effectively addressed under the existing law since Section 51⁴⁰ of the Copyright Act explicitly states that copyright infringement can only be committed by a “person.”⁴¹ In addition to the aforementioned challenges, there is a potential issue with the uniqueness or novelty of the output generated by AI systems. The terms of service of AI platforms acknowledge that the output may not be unique across users for similar questions. Consider a situation where one user claims copyright over a specific output, and then another user independently arrives at the exact same output and also claims copyright over it. If such claims arise on a large scale, it could create complexities. Unlike trademark law, copyright law does not recognize the concept of honest and concurrent use, as it is unlikely for two individuals to come up with the exact same play or book. Unlike typical copyright infringement cases where one party is clearly at fault, the outcome in these circumstances would be difficult to predict. Furthermore, the doctrines of “*scène à faire*” and merger may be applicable to the output generated by AI systems, particularly when the input is a basic question rather than something complex.⁴² These doctrines suggest that such output may not be eligible for copyright protection, as it lacks the necessary originality or expression to qualify for copyright.

The user who claims copyright in the output generated by AI systems faces a challenging situation because the use of such output may not even qualify as fair dealing. When the output generated by AI systems is used in a commercial or public manner, it may not qualify as fair use unless it meets the criteria of being sufficiently transformative. For example, certain outputs obtained when

³⁹ Janhavi Meshram, *Artificial Intelligence Art and Indian Copyright Registration*, MONDAQ, (Nov. 15, 2022), Artificial Intelligence Art And Indian Copyright Registration - Copyright - India (mondaq.com)

⁴⁰ *supra* note 17, § 51.

⁴¹ Rommel Khan, *AI Works – The Future Of Intellectual Property Law - Copyright – India*, (Feb. 20, 2023) MONDAQ, <https://www.mondaq.com/india/copyright/1284668/ai-works--the-future-of-intellectual-property-law>.

⁴² *India recognises AI as author of a copyrighted work*, LEXCAMPUS (Aug. 11, 2021), <https://www.lexcampus.in/india-recognises-ai-as-author-of-a-copyrighted-work/>.