
DEFAMATION WITHOUT A HUMAN AUTHOR: EXAMINING LIABILITY FOR AI-GENERATED CONTENT IN INDIA

Viva Bhusawalia, BA LLB, Bharati Vidyapeeth New Law College, Pune

ABSTRACT

Artificial Intelligence has become a part of everyday life, with tools capable of generating text, images, and other forms of content within seconds. However, these systems can sometimes produce false or misleading information that may harm a person's reputation. This raises an important legal question: can AI commit defamation, and if not, who should be held responsible for such content? This article examines the concept of defamation in the context of generative AI and analyses whether existing Indian laws are equipped to address the challenges posed by AI-generated defamatory statements. It further explores the possible liability of developers, platforms, and users, while highlighting the need for a clearer legal framework to ensure accountability in the age of artificial intelligence.

Keywords: Artificial Intelligence, Defamation, Liability, Generative AI, Indian Law.

I. Introduction

According to the Press information bureau Government of India's "A Roadmap to Viksit Bharat" India is witnessing significant advancements in the field of Artificial Intelligence [AI] with the help of government initiatives that aim towards technological development of the country and provide greater access to computing, research facilities and innovation opportunities. Therefore, AI is increasingly becoming accessible to students, startups and businesses.

However, the rapid growth and adoption of Artificial Intelligence has also raised some severe legal and ethical concerns. One of the most concerning issues is the generation of false or inaccurate statements by AI systems. Generative AI sometimes produce information or response that may appear convincing but are factually incorrect, this phenomenon is referred to as "Hallucination". These can cause reputational harm as responses of such kind may contain false information about an individual, business or public figure. Unlike traditional forms of defamation where a human author can be identified, AI generated content creates uncertainty regarding accountability and liability.

Existing legal frameworks were developed on the premise that defamatory statements originate from human actor. However, when the source of such statement is an autonomous AI, it becomes difficult to determine whether liability should be imposed on the developer, the platform provider, the user or another party.

The present article seeks to examine whether AI can commit defamation and whether the existing legal provisions are capable of addressing such situations. It further explores the topic of liability for AI generated defamatory content and the need for a more coherent legal framework in the age of generative Artificial Intelligence.

The issue becomes even more significant as artificial intelligence continues to be integrated into search engines, customer service platforms, social media applications, and professional workplaces. Information generated by AI is often perceived as reliable due to the sophistication of these systems. Consequently, users may rely on AI-generated responses without independently verifying their accuracy. This increases the risk of false information being disseminated on a large scale and highlights the need to examine whether existing legal principles are capable of addressing the unique challenges posed by AI-generated content.

II. Defamation and AI generated content

Defamation is a legal wrong that arises when false information is communicated to a third person and causes injury to reputation of an individual. The law of defamation protects the reputation of a person which has been recognized as an important aspect personal dignity and social standing. Traditionally defamation consist of a human author that produces a false or incorrect statement with the intention to harm the other persons reputation. However, with the involvement of Generative AI, this conventional understanding has become complicated.

Generative AI systems are capable of producing images, texts, audio and various types of responsive based on the user's prompt. The systems rely on vast amount of training data and statistical patterns. They are trained to provide useful and coherent responses but they are not always accurate. On certain occasions, AI generated information may seem convincing and credible but is entirely false, this is commonly referred as "AI Hallucination".

Several real-life incidents have demonstrated the impact and harm of AI Hallucinations. In 2023, a radio host in the United States alleged that an AI chat box had falsely accused him of financial misconduct by generating inaccurate information of a legal case. Another widely reported incidents is lawyers submitted legal documents containing fictitious judicial decisions generated by AI tools, this highlights the tendency of such systems to provide false information with confidence. These incidents prove that AI Generated inaccuracies are not merely theoretical but practical challenges with potentially serious legal implications.

The possibility of defamatory content being generated by AI raised the questions that the traditional defamation laws were not designed to answer. When a newspaper publishes a defamatory statement, the author and the publisher can be simple to identify, but when AI generates content, there can be multiple authors, including the developers, the company that operates the platform, the user who entered the prompt etc. Among these and many more authors that may be involved, determining who bears the liability is far from straightforward.

III. Who should be liable for AI-generated defamation?

One of the biggest challenges posed by AI-generated defamatory content is determining who should be held legally responsible for the harm caused. Traditional defamation law is built on the assumption that a defamatory statement is made or published by a human being. With

generative AI, however, the situation becomes much more complicated because several parties may be involved in the creation and dissemination of the content.

One possible approach is to hold AI developers liable. Developers design, train, and maintain these systems, and therefore play a significant role in shaping their outputs. It can be argued that they should bear some responsibility when their systems generate harmful or defamatory content. However, imposing liability on developers for every inaccurate statement produced by an AI model may not be practical. Once an AI system is released to the public, it interacts with countless users and responds to an enormous variety of prompts, many of which cannot reasonably be anticipated by its creators.¹

Another perspective focuses on the companies that own and operate AI platforms. These organizations are often in the best position to implement safeguards, monitor problematic outputs, and respond to complaints from affected individuals. If a company becomes aware that its system is generating defamatory content and fails to take reasonable corrective measures, imposing liability may be justified. At the same time, placing unlimited responsibility on AI companies could discourage innovation and slow technological development. Any legal framework must therefore strike a balance between accountability and innovation.²

Liability may also fall on the user who generates or publishes the content. In many situations, AI acts merely as a tool that responds to instructions given by a user. Where a person intentionally uses AI to create false and damaging statements about another individual and then circulates that content, holding the user liable appears to be the most logical approach. The use of AI should not allow individuals to evade responsibility for conduct that would otherwise amount to defamation.

A more controversial question is whether AI itself can be held liable. At present, the answer is largely negative. Existing legal systems recognize liability only for entities that possess legal personality, such as natural persons and corporations. AI systems do not have independent legal status, rights, duties, or the capacity to understand the consequences of their actions. They cannot appear before courts, pay compensation, or be subjected to legal sanctions. For these

¹ Brian Christian, *The Alignment Problem* (2020).

² *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.

reasons, AI cannot currently be treated as a legal person capable of committing defamation.³

In the author's opinion, AI should be viewed as a sophisticated technological tool rather than an independent legal actor. The real challenge is not whether AI should be punished, but rather identifying the human or corporate actors who should bear responsibility when AI-generated content causes reputational harm. As the use of generative AI continues to grow, clearer rules regarding liability will be necessary to ensure that victims have effective legal remedies while allowing innovation to continue.

IV. The Indian legal framework and the need for reform

The rapid advancement of generative artificial intelligence has raised important questions about whether India's existing legal framework is equipped to deal with AI-generated defamatory content. Although laws relating to defamation already exist, they were developed in an era when harmful statements were generally made by identifiable human beings. The emergence of AI-generated content has complicated this traditional understanding and exposed certain shortcomings within the current legal system.

The Bharatiya Nyaya Sanhita, 2023 continues to recognize defamation as an offence and provides legal protection against harm caused to a person's reputation. Under Section 356, a person may be held liable for making or publishing statements that are likely to damage another individual's reputation.⁴ While the provision is broad enough to apply to content circulated through digital platforms, it does not specifically address situations where defamatory material is produced by an AI system. This creates uncertainty when attempting to identify the party responsible for the harm caused.

Certain provisions of the Information Technology Act, 2000 may also be relevant in this context. Section 79 offers protection to intermediaries for third-party content, provided they comply with prescribed due diligence requirements.⁵ The Supreme Court's decision in *Shreya Singhal v. Union of India* further clarified the responsibilities of online intermediaries and the regulation of online speech. However, AI platforms differ from traditional intermediaries because they do more than merely host content—they actively generate it. As a result, existing

³ Ugo Pagallo, *The Laws of Robots* (2013).

⁴ Bharatiya Nyaya Sanhita, 2023.

⁵ Information Technology Act, 2000.

legal principles do not provide a complete answer to questions of liability arising from AI-generated statements.

A major gap in the present framework is the absence of clear rules regarding accountability. Traditional defamation law assumes that a defamatory statement originates from a human speaker or publisher. In contrast, AI-generated content may involve multiple actors, including developers, platform providers, and end users. The law currently offers limited guidance on how responsibility should be distributed among these parties.

To address these concerns, certain reforms should be considered. First, AI platforms should be required to provide clear disclosures informing users that the generated content may contain inaccuracies or misleading information. Such transparency would encourage users to verify information before relying upon it.

Secondly, lawmakers should develop clearer liability standards to determine the circumstances in which developers, companies, or users may be held responsible for defamatory AI-generated outputs. Greater clarity would benefit both victims seeking remedies and organizations deploying AI technologies.

Finally, an efficient correction mechanism should be introduced. Individuals affected by inaccurate AI-generated information should have access to a straightforward process through which false content can be challenged, corrected, or removed. Such safeguards would help reduce reputational harm while promoting the responsible use of artificial intelligence.

As AI technology continues to evolve, the legal framework must also adapt. A balanced approach that protects individual reputation without stifling innovation will be essential in addressing the challenges posed by AI-generated defamation.

V. Conclusion

The rise of generative artificial intelligence has created new challenges for traditional defamation law. AI systems are capable of producing content that appears accurate and reliable, yet they may occasionally generate false statements that can damage an individual's reputation. As AI-generated content becomes increasingly common, the possibility of reputational harm caused by such systems can no longer be viewed as a purely theoretical concern.

At present, however, artificial intelligence cannot be regarded as a legal person capable of bearing responsibility for its actions. Unlike natural persons or corporations, AI systems do not possess legal personality, intent, rights, or duties. Consequently, liability for defamatory content generated through AI must ultimately be traced to the human actors involved in its development, operation, or use.

The study highlights that existing Indian laws relating to defamation and online intermediaries were not designed with generative AI in mind. While these laws may provide limited guidance, they do not adequately address the unique questions of accountability raised by autonomous content generation. As a result, significant uncertainty remains regarding who should be held liable when AI-generated statements cause reputational harm.

To address these challenges, India should consider developing clearer legal standards governing AI-related liability. A balanced regulatory framework that promotes accountability, transparency, and effective remedies for victims will be essential to ensure that technological innovation progresses without compromising the protection of individual reputation.