The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2024

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

LEGAL LIABILITY IN THE AGE OF AI-GENERATED CONTENT

ABSTRACT

Numerous legal issues have emerged as a result of the widespread use of artificial intelligence (AI) technology, most notably those pertaining to the responsibility for content produced by AI systems. This study examines the complex web of moral and legal issues related to artificial intelligence-generated content that contains offensive or unlawful themes. This paper explores the potential liability of AI developers, content platforms, and users in the event that AI-generated content violates Indian laws pertaining to defamation, hate speech, incitement, and other offenses. It does this by looking at pertinent legal frameworks, case studies, and community perspectives. The study assesses potential defenses and mitigation techniques while analyzing the roles and duties of key stakeholders based on precedents and actual instances. The paper also examines the moral ramifications of AI-generated content and provides guidance to researchers, business interests, policymakers, and regulators on how to negotiate the murky waters of AI liability in India. This study seeks to add to the current conversation on AI policy and online content governance in the Indian context by bringing attention to this urgent topic.

Keywords: Artificial Intelligence, Liability, Hallucination, AI-generated contents, Legal position, Policy drafting

INTRODUCTION

Artificial intelligence (AI) technology has completely changed a number of aspects of contemporary life, including communication, entertainment, and even business and healthcare. But while AI develops quickly, it also presents a number of moral and legal issues, especially when it comes to creating and sharing material. The emergence of AI-generated material poses

important problems about liability and accountability in the event of information that is illegal or seditious in character in the Indian setting, where the digital landscape is fast expanding.

This study aims to investigate the complex legal environment around artificial intelligence (AI)-generated content that contains particular phrases that could incite sedition or break other Indian laws. With the development of AI technologies, such as deep learning algorithms and natural language processing, computers are now able to produce content on their own or with little assistance from humans. This offers great opportunities for creativity and innovation, but it also carries a number of serious hazards, especially when it comes to information that can be illegal in India.

Analyzing the possible culpability of different parties engaged in the production, distribution, and control of AI-generated content in India is the main goal of the present paper. This encompasses content platforms, users, regulators, law enforcement, and AI developers. The objective of this article is to clarify the intricate relationship between legal, ethical, and technological elements influencing the liability environment in India through an analysis of pertinent legal frameworks, case studies, and user thoughts.

This paper will also examine the ethical consequences of AI-generated content, taking into account issues with censorship, free speech, and harm to society. This study aims to offer insights into how Indian policymakers, regulators, industry stakeholders, and researchers might negotiate the obstacles faced by AI-generated content with seditious or unlawful keywords by critically analyzing current literature, legal precedents, and real-world experiences.

In the end, this study article attempts to add to the current conversation in India on AI regulation and online content governance. This study attempts to inform and enable stakeholders to confront the complex difficulties provided by AI-generated material in the Indian legal landscape by identifying important legal issues, examining pertinent case studies, and making recommendations for future action.

AN OVERVIEW ON AI

One of the contemporary trends is the rapid advancement of human scientific activities. New technologies are crashing into our lives every day. Tim Urban, founder of the well-known website Wait but why, which explores various issues, including artificial intelligence (AI), claims that AI is not just an important topic but also by far the most crucial component of our future. AI offers opportunities to complement and supplement human intelligence and enhance the way people live and work. Intelligent machines are enabling high-level cognitive processes like thinking, perceiving, learning, problem solving, and decision making. These advancements

are paired with advances in data collection and aggregation, analytics, and computer processing power[1]. The vast field of artificial intelligence (AI) encompasses a range of intelligence and capability levels. Artificial intelligence is divided into three levels: ANI, AGI, and ASI.

The initial level capable of achieving a decade in a single domain is called Artificial Narrow Intelligence, or ANI. Artificial narrow Intelligence (ANI) is a fundamental concept that describes AI systems built to accomplish particular tasks inside a small domain. Though its capabilities are restricted to the specific domain it was developed for, ANI excels at problem resolution and task execution within its prescribed scope. ANI is represented, for instance, by AI systems designed to play chess or examine medical pictures. Although these systems are able to accomplish jobs far more efficiently than people, they are not flexible enough to apply their knowledge or abilities outside of their own field.

Artificial General intellect (AGI) is defined as AI that approaches and surpasses human intellect, which means it can reason, plan, solve problems, think abstractly, understand complicated concepts, pick things up rapidly, and learn from mistakes[2]. Artificial General Intelligence (AGI), is a major development over Artificial Neural Networks (ANI). Artificial intelligence (AGI) is defined as AI systems that are similar to human intelligence in that they can think, interpret, and learn in a variety of disciplines. AGI may solve problems, reason abstractly, and learn from experience in a way that goes beyond boundaries, in contrast to ANI, which functions within a predetermined domain. AGI has the capacity to execute a large variety of cognitive activities at a level that is on par with or even higher than that of humans. Even while AGI is still mostly theoretical, its advancement marks a significant turning point in the search for highly intelligent machines.

Artificial Super Intelligence, or ASI, is intelligence that surpasses the best human brain in almost every area, including social skills, general knowledge, and creative problem-solving[3]. Artificial Super Intelligence (ASI), the highest level of machine intelligence that surpasses human intellect in almost every domain, is at the pinnacle of the AI hierarchy. ASI represents intelligence that, in a variety of domains, such as social interaction, creativity, and general knowledge, not only equals but surpasses the cognitive prowess of the most brilliant human minds. ASI predicts a revolutionary future in which machines will be endowed with not only intelligence but also a degree of creativity and understanding that will be incomprehensible to

[1] National Strategy for Artificial Intelligence by NITI Aayog, June 2018, p.no. 5

[2] Linda S. Gottfredson Mainstream Science on Intelligence : An Editorial With 52 Signatories, History and Bibliography / Linda S. Gottfredson. - Ablex Publishing Corporation. 1997.

[3] Nick Bostrom How long before superintelligence? / Nick Bostrom. - Linguistic and Philosophical Investigations, 2006. - pp. 11-30.

humans. Although ASI is yet theoretical and speculative, it has enormous and far-reaching potential effects on technology, society, and the human condition.

REAL LIFE INCIDENTS OF AI CREATING DISINFORMATIONS

Researchers from the Regulation, Evaluation, and Governance Lab at Stanford University and the Institute for Human-Centered AI conducted a study that produced alarming results about the dependability of large language models (LLMs) in legal contexts. The study found that well-known LLMs, including ChatGPT's GPT-3.5, Google's PaLM 2, and Meta's Llama 2, showed notable rates of hallucination and gave false knowledge when asked direct and verified questions concerning federal court cases. In particular, GPT-3.5 experienced hallucinations 69% of the time, but PaLM 2 and Llama 2 provided incorrect information 88% and 72% of the time, respectively. Furthermore, the models had trouble answering increasingly complicated legal problems. Especially when it came to case law from subordinate courts like district courts, they frequently misidentified the major issues raised by a case or its central holdings. The study also found that these LLMs tended to overestimate their confidence in their responses and often failed to confront erroneous premises inside legal issues, raising serious concerns regarding their fitness and dependability for use in legal settings[4].

After using Google Bard, which was operating on PaLM 2 until recently, Michael Cohen, the former president Trump's personal attorney and fixer, also provided his lawyer with fictitious case citations[5].

Chief Justice John Roberts cautioned about the possible negative effects of utilizing AI in the legal profession while also hinting that the technology may have a big impact on judicial work in the future. "Any use of AI requires caution and humility," he said. In his annual year-end report[6].

Gwinnett County, Georgia, Superior Court Judge Tracie Cason says businesses should be aware that generative AI "hallucinations" could expose them to liability in January 2024[7].

In June 2023, after a journalist asked ChatGPT to summarize a case concerning the Second Amendment Foundation, Mark Walters launched a defamation action against OpenAI.

---

[4] Dahl, Matthew, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. "Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive." Jan. 11, 2024.

[5] Weiser, Benjamin and Jonah E. Bromwich. "Michael Cohen Used Artificial Intelligence in Feeding Lawyer Bogus Cases." The New York Times, Dec. 29, 2023.

[6] Year-End Report on the Federal Judiciary 2023, p. 6.

[7] Ballard Spahr. "Judge Denies Motion to Dismiss AI Defamation Suit." By Charley F. Brown and Jonathan P. Hummel. January 24, 2024.

ChatGPT replied that Mr. Walters had embezzled funds from the Foundation and even created fake sections from the case[8];

In April 2023, in response to a request for instances of academic sexual harassment at US law schools, ChatGPT produced a bogus report claiming that Jonathan Turley, a professor, had been the subject of sexual misconduct charges, and even provided a phony Washington Post story to bolster the claims[9];

In March 2023, an Australian mayor threatened to file a defamation lawsuit against OpenAI when the chatbot, ChatGPT, falsely stated that Brian Hood had been found guilty in the Securency bribery case, despite the fact that he was the whistleblower[10];

In 2016 Microsoft introduced Tay, a chatbot on Twitter. Tay was created to mimic human speech and learn from user interactions. But after being tricked by users, Tay started tweeting provocative and abusive content just hours after it launched. Tay had to be shut down, and Microsoft had to apologize. This event brought up concerns around accountability and culpability for content created by AI, even if it was not a legal case.

LEGAL POSITION OF AI

The individual (natural person or legal organization) for whom a computer was intended bears ultimate responsibility for any communication produced by the device, per Article 12 of the United Nations Convention on the Use of Electronic Communications in International Contracts[11]. This point of view adheres to the fundamental principle that, as a tool lacks free will, its creator is accountable for the outcomes that can be achieved with it. The Information Technology Act, 2000's provisions[12] grant legal validity to electronic contracts in India. Section 11(c)[13] makes it clear that an information system can be configured on behalf of a human being when it states, "by an information system programmed by or on behalf of the originator to operate automatically."

The Companies Act of 2013 permits the establishment of "One Person Companies," and it can be modified to permit the establishment of "Zero Person Companies" and other comparable entity structures that would enable the functional personhood of an AI system. Legally

[8] Thaler, Shannon. "ChatGPT 'hallucination' falsely said radio host embezzled money, suit says." New York Post. Published June 7, 2023.

[9] Cost, Ben. "ChatGPT smeared me with false sexual harassment charges: law professor." New York Post. Published April 7, 2023. Updated April 7, 2023.

[10] ABonyhady, Nick. "Australian whistleblower to test whether ChatGPT can be sued for lying." The Sydney Morning Herald, April 5, 2023.

[11] https://treaties.un.org/doc/source/RecentTexts/X-18_english.pdf.

[12] Section 10A, Information Technology Act, 2000.

[13] Section 11(c), Information Technology Act, 2000.

speaking, artificial intelligence (AI) could be considered a "person" since, in contrast to corporations, AI is truly autonomous—that is, until a certain point, when its programmers relinquish control and it acts entirely of its own volition. Thus the contracts created through the interaction between an automated system and a human being are recognized as legitimate and enforceable.

i. Theory of *Respondeat Superior*

The master-servant rule, or respondeat superior concept, had its beginnings in ancient Rome and was applied for the first time in England in the 1698 case of Jones v. Hart. It declares that a party is accountable for the deeds of their delegates. It's crucial to remember, though, that AI and slaves are both regarded as objects of law and do not have the ability to file lawsuits since they are not acknowledged as legal subjects. Nevertheless, there might be comparisons made between AI's legal standing and that of slaves in terms of accountability for their deeds. The owner, AI creator, or legal body on whose behalf the AI acts may be held accountable for harm brought about by AI actions, in the same way that slaveholders were held accountable for the wrongdoings of their slaves. According to this theory, the person or thing in charge of the AI may be held liable for any wrongdoings carried out by it, much like the head of a family in charge of slaves.

ii. Doctrine of Vicarious Liability

The idea of vicarious responsibility, which has its origins in agency law, holds people responsible for the wrongdoings of others because of their association with the wrongdoer. Owners or users are usually accountable for the behavior of the AI systems under their control in the context of AI. In this connection, AI functions as the principal's agent, much like in an agency. As per international agreements and e-commerce regulations, the principal bears ultimate responsibility for any communications or transactions started by AI systems[14]. Courts may need to modify agency law to establish accountability for injuries caused by AI as the technology grows more independent and capable of making decisions on its own. Potential legal frameworks for resolving AI-related liabilities can be gleaned from precedents from incidents involving automated technology, such as the Ashley Madison data breach[15]. New methods of approaching AI jurisprudence are, however, required as AI develops to make

[14] Paulius, C., Grigien, J., & Sirbikytė, G. (2015).
[15] 23 F. Supp. 3d 1378, 1380 (JPML 2015).

judgments outside of its programming and concerns the application of conventional agency principles surface.

iii. Criminal Liability

Criminal responsibility must be established by both the criminal act and the criminal purpose, or Actus Reus and Mens Rea, respectively. Liability cannot be shown without the occurrence of both. Actus reus is usually illustrated by deeds or omissions; occasionally, more factual elements, including results and circumstances, are required. From knowing to intent or explicit aim, mens rea spans a range of mental states, with criminal negligence or recklessness reflecting lower degrees. According to Gabriel Hallevy's[16] proposal, three liability models—perpetration-by-another, natural-probable consequence, and direct responsibility—can assign criminal accountability to AI entities. These models reflect the difficulties in determining criminal responsibility in an increasingly automated society and provide frameworks for evaluating AI's role in criminal activities.

iv. Liability under Consumer Act

AI corporations need to be aware of the consequences of the recently passed Consumer Protection Act of 2019, especially Section 2(47), which defines "Unfair Trade Practice." This definition includes deceptive information and fabrications, but it makes no allowances for AI algorithms that use automated decision-making. Therefore, companies that develop AI or those that use commercial AI solutions can be vulnerable to consumer protection complaints, which are now easily submitted online from the comfort of one's home. The Act includes the disclosure of personal information provided in confidence by the consumer as part of the definition of unfair commercial practices, unless disclosures are required by law or serve the public interest. In order to avoid potential consequences from the law, AI firms must prioritize openness and responsibility in their AI applications and assure compliance with consumer protection rules.

v. Common Enterprise Liability

It can be difficult to assign blame when defects or hacks occur in automated systems; possible suspects include automakers and individual component designers. Existing models of causal responsibility perform best when computer operations can be linked to human design; in other

[16] Hallevy, G. (2010)

cases, they are not very effective. Imposing strict liability on automated system creators is one suggested remedy. Alternatively, a group of people involved in the design and production of AI could share guilt if the Common Enterprise Doctrine were applied to a recently enacted severe liability system. Legal expert David C. Vladeck proposed this method, which makes all members of a connected group accountable for each other's acts even in the absence of direct cooperation. As demonstrated in cases like *SEC v. R.G. Reynolds Enters., Inc*.[17], *FTC v. Tax Club, Inc.*[18], and *FTC v. Network Services Depot, Inc*.[19], regulatory bodies such as the Federal Trade Commission frequently employ comparable doctrines, such as the "common business" doctrine, to pursue joint and multiple responsibility among interconnected businesses engaged in fraudulent activities.

vi. Defamation

If done in a way that is appropriate for the situation, "publishing," or telling third parties something incorrect about someone else that damages their reputation, is considered defamation. In their current state, generative models do not ensure that the material they produce will be true to the training data. They sample subsequent words or symbols in a probabilistic manner by nature, and they are prone to making unfounded allegations. It has been discovered that the majority of us will be the subject of deceptive speech based on recent popular models. This implies that businesses using generative models, which frequently produce incorrect statements, may be sued. Eugene Volokh goes into great detail about this possibility in his work[20].

WHO ARE LIABLE?

Various parties involved in the development, implementation, and use of the AI system can be held accountable for the output of generative AI, especially when it is factually erroneous or harmful. If an AI system is designed, coded, or produced by a developer, they may be held accountable for producing a product that has a known tendency to provide hazardous or deceptive results. But whether an AI system is created internally or externally, the deployer—the person who uses it in a specific situation—is equally accountable for making it available to

[17] S.E.C. v. R.G. Reynolds Enterprises, Inc., 952 F.2d 1125.
[18] FTC v. Tax Club, Inc. 994 F. Supp. 2d 461, 471 (S.D.N.Y. 2014).
[19] FTC v. Network Services Depot, Inc. is 617 F.3d 1127, 1138 (9th Cir. 2010).
[20] Eugene Volokh, Large Libel Models? Liability for AI Output, 3 J. Free Speech L. 489, 555 (2023).

the general public. In addition, the person who initiates the prompt and receives the response from the AI system—known as the prompter—may also be held responsible, particularly if they intentionally create prompts that trick the system into producing offensive information. The action for which a developer, deployer, or user is liable varies. Liability for the user may result from releasing the AI-generated output into the public domain or from purposefully encouraging harmful or inaccurate information. If they were informed about the limitations of the AI by the developer or deployer, they may be held accountable for hate speech or defamation just like if they had produced the content themselves. Liability for the developer and deployer arises from creating or making available a product that has a recognized tendency to produce output that is erroneous or harmful.

MORAL ASPECTS

It's critical to acknowledge the necessity of regulation and control for AI content creation. Establishing moral standards and legislative frameworks to control the proper application of AI technology is mostly the responsibility of governments and international organizations. This could entail creating guidelines for accountability, fairness, and openness in the creation of AI content as well as systems for compliance and enforcement.

Secondly, the broad use of AI in content creation will change how information is created, shared, and consumed, which will have a significant impact on society. Comprehending these ramifications is crucial for foreseeing and managing possible obstacles including employment displacement, modifications in communication patterns, and adjustments in power structures. The broader societal effects of AI-driven content creation should be taken into account by ethical norms, which also aim to make sure that technology advances human welfare and serves the public interest.

Ultimately, since bias in AI-generated content has the potential to encourage discriminatory actions and maintain current disparities, it is a serious problem. It's critical to address bias in the data sets used to train AI algorithms and to put policies in place to identify and fix biased outputs in order to reduce bias. This could entail creating algorithms that put equity and justice first, conducting bias audits, and diversifying the training set. Prioritizing the reduction of bias and encouraging the creation of AI systems that are just, inclusive, and considerate of human rights should be the main goals of ethical guidelines.

CONCLUSION

In light of the pervasiveness of biases in the data that is currently available and the opaque nature of many algorithms, the drive for justice and transparency in AI systems is imperative. Proactively identifying and mitigating biases is necessary for addressing them, and opening the "Black Box" to decision-making processes improves transparency. Achieving this balance is difficult and calls for cooperative study and discussion to guarantee that AI systems are reliable, responsible, and free from unforeseen repercussions[21]. In June 2018, the National Strategy for Artificial Intelligence, a policy study published by NITI Aayog in India, examined the significance of AI across many industries. It was also suggested in Budget 2019 to start a national AI initiative. Despite all of these advancements in technology, the nation has not yet developed comprehensive legislation to control this expanding sector of this industry. The Indian government needs to respond to these kinds of situations as soon as possible.

REFERENCES

1. ABonyhady, Nick. "Australian whistleblower to test whether ChatGPT can be sued for lying." The Sydney Morning Herald, April 5, 2023.
2. Ballard Spahr. "Judge Denies Motion to Dismiss AI Defamation Suit." By Charley F. Brown and Jonathan P. Hummel. January 24, 2024.
3. Bostrom, Nick. "How long before superintelligence?" Linguistic and Philosophical Investigations (2006), pp. 11-30.
4. Cost, Ben. "ChatGPT smeared me with false sexual harassment charges: law professor." New York Post, April 7, 2023 (updated April 7, 2023).
5. Dahl, Matthew, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. "Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive." Jan. 11, 2024.
6. Eugene Volokh. "Large Libel Models? Liability for AI Output." 3 J. Free Speech L. 489, 555 (2023).
7. FTC v. Network Services Depot, Inc., 617 F.3d 1127, 1138 (9th Cir. 2010).
8. FTC v. Tax Club, Inc., 994 F. Supp. 2d 461, 471 (S.D.N.Y. 2014).
9. Hallevy, G. (2010).
10. Gottfredson, Linda S. "Mainstream Science on Intelligence: An Editorial With 52 Signatories, History and Bibliography." Ablex Publishing Corporation (1997).
11. National Strategy for Artificial Intelligence by NITI Aayog, June 2018, p. 5.

[21] National Strategy for Artificial Intelligence by NITI Aayog , June 2018, p.no. 85 & 86

12. National Strategy for Artificial Intelligence by NITI Aayog, June 2018, pp. 85-86.
13. Paulius, C., Grigien, J., & Sirbikytė, G. (2015).
14. Section 10A, Information Technology Act, 2000.
15. Section 11(c), Information Technology Act, 2000.
16. S.E.C. v. R.G. Reynolds Enterprises, Inc., 952 F.2d 1125.
17. Thaler, Shannon. "ChatGPT 'hallucination' falsely said radio host embezzled money, suit says." New York Post, June 7, 2023.
18. Weiser, Benjamin and Jonah E. Bromwich. "Michael Cohen Used Artificial Intelligence in Feeding Lawyer Bogus Cases." The New York Times, Dec. 29, 2023.
19. Year-End Report on the Federal Judiciary 2023, p. 6.
20. https://treaties.un.org/doc/source/RecentTexts/X-18_english.pdf.