WHEN AI GETS IT WRONG: WHO IS LEGALLY RESPONSIBLE FOR AI HALLUCINATIONS IN INDIA?

Puspaak Ray, Law Student, Maharishi Markandeshwar University, Haryana

ABSTRACT

Artificial intelligence (AI) has rapidly transitioned from a specialized technological tool to an integral component of everyday decision-making processes. Its growing use in areas such as legal research, documentation, and advisory functions has raised significant concerns regarding the reliability of its outputs. One of the most pressing issues is the phenomenon of "AI hallucination," wherein AI systems generate factually incorrect or fabricated information while presenting it with apparent confidence and authority.This paper examines the legal implications of AI hallucinations within the Indian context, focusing on the challenges they pose to traditional doctrines of liability. In particular, it explores the difficulty of applying the concept of mens rea to AI systems, which lack consciousness, intent, and moral agency. The absence of a "guilty mind" complicates the attribution of criminal liability and necessitates a shift in focus toward human actors involved in the design, deployment, and use of such technologies. The paper further analyses the applicability of existing Indian legal frameworks, including the Information Technology Act, 2000, the Consumer Protection Act, 2019, and the Digital Personal Data Protection Act, 2023, in addressing harms caused by AI-generated misinformation. It highlights the regulatory gaps and accountability challenges that arise due to the evolving nature of AI systems.Finally, the study proposes a structured approach toward liability, emphasizing the need for clearer legal standards, enhanced transparency, and greater user awareness. It argues that while AI cannot be held liable in itself, a balanced framework assigning responsibility across stakeholders is essential to ensure accountability in an increasingly automated world.

Key Words: - *Artificial Intelligence, AI Hallucinations, Legal Liability, Mens Rea, Consumer Protection, India.*

I. INTRODUCTION

Artificial intelligence is no longer a futuristic concept reserved for tech experts; it has quietly become a part of everyday decision-making. From answering legal queries to drafting documents, AI tools are increasingly being relied upon for tasks that once required human judgment. However, this growing dependence comes with a hidden risk: these systems can sometimes generate information that is entirely incorrect, yet presented with complete confidence. This phenomenon, commonly referred to as "AI hallucination," has begun to raise serious concerns. Unlike human errors, where intent or negligence can be examined, AI-generated mistakes exist in a grey area. If an individual suffers harm after relying on incorrect information provided by an AI system, it becomes difficult to determine where responsibility lies. Can the machine itself be blamed, or does liability shift to the humans behind its creation and use?

In the Indian legal context, this question becomes even more complex due to the absence of a specific regulatory framework addressing AI accountability. As technology continues to evolve faster than the law, it is crucial to examine how existing legal principles can respond to such challenges. This article seeks to explore the issue of AI hallucinations through the lens of liability and accountability, and to identify who should ultimately bear responsibility when artificial intelligence gets it wrong.

II. WHAT ARE AI HALLUCINATIONS?

Artificial intelligence systems, particularly those based on machine learning and large language models, function by identifying patterns in vast amounts of data and generating responses based on probability rather than understanding.[1] While this allows them to produce quick and seemingly coherent answers, it also creates room for significant error.6 Unlike human mistakes, these outputs are not the result of carelessness or deliberate misrepresentation. Instead, they arise from limitations in training data, gaps in contextual understanding, or the model's tendency to "fill in" missing information with plausible-sounding content.For instance, an AI tool may cite judicial decisions that do not exist, misinterpret statutory provisions, or provide outdated legal advice while presenting it as current. To an average user, especially one without legal expertise,

[1] Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach* (3rd edn., Pearson 2010).

such responses can appear entirely reliable. This makes AI hallucinations particularly dangerous, as they blur the line between accurate information and convincing misinformation.The key issue lies in the illusion of certainty. AI systems do not possess the ability to verify truth in the way humans do; they merely predict what a correct answer might look like. As a result, their outputs can carry an unwarranted sense of authority, increasing the risk of misuse and harm when relied upon without independent verification

III. THE CORE LEGAL PROBLEM: ABSENCE OF MENS REA

One of the foundational principles of criminal law is that liability is not imposed merely for committing a wrongful act, but for doing so with a guilty mind, known as *mens rea*.[2] This mental element includes intention, knowledge, recklessness, or negligence, and it plays a crucial role in determining culpability. Courts have repeatedly emphasized that without *mens rea*, criminal liability is generally difficult to establish. However, applying this principle to artificial intelligence creates an immediate conceptual problem. AI systems do not possess consciousness, awareness, or intent. They do not "decide" to produce incorrect information, nor do they understand the consequences of their outputs.

This distinction is critical. In human conduct, a false statement can amount to deception or fraud if it is made knowingly or with intent to mislead. In contrast, an AI-generated falsehood lacks this essential mental element. The system is not lying in the legal sense; it is merely producing an incorrect output based on its programming and data patterns.This absence of *mens rea* makes it difficult to attribute direct liability to AI systems under traditional criminal law frameworks. If there is no guilty mind, the question arises whether there can be any guilt at all. As a result, the focus of legal responsibility shifts away from the machine itself and toward the human actors involved in its design, deployment, and use.This shift highlights a deeper challenge: existing legal doctrines are built around human behaviour and moral agency. When harm is caused by entities that lack both, the law must either adapt its principles or risk leaving significant gaps in accountability.

[2] R v Prince (1875) LR 2 CCR 154.

IV. WHO CAN BE HELD RESPONSIBLE:

Since artificial intelligence systems cannot be held liable in the traditional legal sense due to the absence of intent or consciousness, the question of responsibility must shift to the human actors connected to their functioning. Determining liability in cases of AI hallucinations is therefore less about blaming the machine and more about identifying where human accountability lies within the system.

1. Developers and Designers

Developers play a foundational role in shaping how AI systems function. If an AI tool produces harmful or misleading outputs due to poor design, inadequate testing, or failure to address foreseeable risks, liability may arise under the principles of negligence.[3] The law imposes a duty of care on individuals to ensure that their actions do not cause harm to others. In the context of AI, this duty extends to designing systems that are reasonably safe and reliable.

2. Companies and Platforms

Organizations that deploy AI systems for public or commercial use may also bear significant responsibility. These entities control how the technology is presented, marketed, and integrated into user-facing services. If an AI platform is portrayed as a reliable source of information without adequate disclaimers or verification mechanisms, users may reasonably rely on its outputs. If users rely on AI outputs due to misleading representations or lack of adequate disclaimers, liability may arise under consumer protection laws for deficiency in service.[4]

3. Users and End-Operators

While much of the focus is placed on developers and companies, users are not entirely exempt from responsibility. In situations where individuals rely blindly on AI-generated information without exercising basic caution or verification, their conduct may amount to negligence.

[3] Donoghue v Stevenson [1932] AC 562 (HL).
[4] Consumer Protection Act, 2019.

V. THE INDIAN LEGAL POSITION

India does not yet have a comprehensive legal framework specifically addressing liability for artificial intelligence or the consequences of AI-generated errors such as hallucinations. However, certain existing laws provide a partial foundation for addressing these issues, even if they were not originally designed with AI in mind.

The Information Technology Act, 2000 governs digital intermediaries and online platforms.[5] Under this framework, intermediaries are generally granted limited liability for third-party content, provided they do not have actual knowledge of unlawful material or fail to act upon receiving such knowledge. However, AI-generated content complicates this position. Unlike traditional intermediaries that host user-generated content, AI platforms themselves generate the information in question, making it difficult to clearly categorize their role.

The Consumer Protection Act, 2019 may also become relevant where AI tools are offered as services to users. If an AI system provides misleading or inaccurate information that results in harm, it could be argued that there is a "deficiency in service."[6] This is particularly important in cases where platforms present their AI systems as reliable or authoritative without clearly communicating their limitations.

Further, the Digital Personal Data Protection Act, 2023 reflects the growing emphasis on responsible data usage in India.[7] Since AI systems rely heavily on data for training and functioning, issues such as data accuracy, consent, and misuse become indirectly linked to the reliability of AI outputs.

Judicial interpretation also provides guidance on how technology should be approached within the constitutional framework. In Justice K.S. Puttaswamy v. Union of India, the Supreme Court recognized the right to privacy as a fundamental right and emphasized that technological advancements must operate within constitutional limits.[8] This principle suggests that any future regulation of AI in India must balance innovation with the protection of individual rights. This

[5] Information Technology Act, 2000, s. 79
[6] Consumer Protection Act, 2019, s. 2(11).
[7] Digital Personal Data Protection Act, 2023.
[8] Justice K.S. Puttaswamy v Union of India (2017) 10 SCC 1.

ambiguity indicates that existing legal frameworks, while adaptable, are not sufficiently equipped to address the autonomous and generative nature of modern AI systems.

VI. RISKS AND CONCERNS

AI hallucinations pose significant legal and societal risks. One of the primary concerns is the spread of misinformation, which may lead to legal, financial, or reputational harm when users rely on inaccurate outputs. Additionally, the authoritative tone of AI systems often creates an illusion of reliability, resulting in over-dependence and reduced critical evaluation by users.Another major issue is the lack of transparency in AI functioning. Many systems operate as "black boxes," making it difficult to trace the reasoning behind a particular output. This lack of explainability complicates the process of assigning liability when harm occurs. Furthermore, the involvement of multiple stakeholders, including developers, deployers, and users, creates ambiguity in accountability, leading to potential gaps in legal enforcement.

VII. THE WAY FORWARD

Addressing the challenges posed by AI hallucinations requires a structured and forward-looking legal approach. Firstly, there is a need for clearly defined liability frameworks that allocate responsibility among developers, companies, and users based on their level of control and involvement. Secondly, regulatory standards must emphasize transparency and explainability to ensure that AI outputs can be meaningfully assessed.

Further, developers and platforms should be subject to a heightened duty of care, requiring regular system audits, improved data quality, and the incorporation of safeguards to minimize harmful outputs. User awareness also plays a crucial role, as individuals must be encouraged to verify AI-generated information rather than rely on it blindly.

Finally, India must consider the development of a dedicated regulatory framework for artificial intelligence that aligns with constitutional principles while promoting innovation. Such a framework would help bridge existing gaps and ensure accountability in the use of emerging technologies.

VIII. CONCLUSION

Artificial intelligence has introduced a new layer of complexity into legal systems by challenging traditional notions of responsibility and intent. AI hallucinations, in particular, demonstrate how harm can arise without any human-like intention behind it. This disrupts the conventional framework of liability, which is largely built on the presence of *mens rea*.While AI systems cannot be held liable in the legal sense due to the absence of consciousness and intent, the consequences of their outputs cannot be ignored. The law, therefore, must shift its focus from the machine itself to the human actors involved in its creation, deployment, and use. Responsibility must be carefully distributed based on control, knowledge, and the ability to prevent harm.In the Indian context, the absence of a dedicated legal framework highlights the need for reform. Existing laws provide limited guidance, but they are not fully equipped to deal with the unique challenges posed by AI-generated misinformation. Moving forward, a balanced approach that combines regulatory clarity, technological accountability, and user awareness will be essential.

Ultimately, as artificial intelligence continues to evolve, the legal system must ensure that efficiency does not come at the cost of accountability. The question is not whether AI can get things wrong it already does but whether the law is prepared to respond when it does.