VOLUME 6 | ISSUE 4

INTERNATIONAL JOURNAL OF ADVANCED LEGAL RESEARCH

HALLUCINATING JUSTICE: RETHINKING PROFESSIONAL AND DEVELOPER LIABILITY FOR GENERATIVE AI ERRORS IN LEGAL PRACTICE AND LAW

- Aadhya Bhardwaj[1]

ABSTRACT

The increasing reliance on generative artificial intelligence in legal practice has complicated the attribution of liability when hallucinated outputs generate professional, financial, or reputational harm. Existing approaches predominantly impose responsibility upon lawyers through established duties of competence and verification or examine the technical limitations of generative models in isolation, offering limited guidance on how liability should be allocated between legal practitioners and AI developers. Employing a doctrinal and comparative methodology, this paper analyses professional responsibility principles, negligence doctrines, product liability theories, and emerging AI governance frameworks. It suggests that dual models of exclusive lawyer liability and developer immunity inadequately captured the distributed nature of algorithmic risk and suggests a Three-Factor Liability Test, grounded in foreseeability, control, and the relative capacity to prevent harm, as a principle and administrable standard for allocating responsibility in AI-assisted legal services

Keywords: Generative Artificial Intelligence, AI Hallucinations, Professional Liability, Developer Liability, Product Liability, Algorithmic Risk, Legal Ethics, Shared Liability.

INTRODUCTION

Generative artificial intelligence known as GenAI has exposed a growing tension between traditional liability doctrines and the distributed risks generated by probabilistic decision support systems. This tension is particularly lined within legal practice where professional competence, adjudicative integrity, and public confidence in administration of justice depends upon precision, verifiability, and fidelity to authoritative sources. By augmenting legal research, drafting, document review, and case analysis, GenAI offers considerable gains

[1] Student at Guru Gobind Singh Indraprastha University, New Delhi

For general queries or to submit your research for publication, kindly email us at ijalr.editorial@gmail.com

https://www.ijalr.in/

in efficiency and accessibility. Yet its increasing integration into legal decision making raises a fundamental challenge that how technological innovation may be accommodated without compromising the standards of accuracy and accountability that underpin legal services.

Despite their utility, generative AI systems remain susceptible to producing inaccurate or entirely fabricated outputs commonly described as "hallucinations." In legal settings, such errors may manifest as fictitious judicial authorities, erroneous citations, non-existent statutory provisions, or flawed legal analyses presented with persuasive coherence. Their incorporation into pleadings, legal memoranda, or client advisories extends beyond technological fallibility, implicating lawyers' professional obligations, the integrity of adjudicative processes, and the attribution of responsibility where reliance upon AI-generated misinformation causes legal, financial, or reputational harm.

Although existing scholarship has extensively examined AI hallucinations, legal ethics, and emerging models of AI governance, comparatively limited attention has been devoted to the jurisprudential problem of allocating liability where foreseeable hallucinations produce harm within legal services. Existing approaches frequently compartmentalise professional negligence and developer accountability, thereby obscuring the distributed nature of algorithmic risk. This paper contends that binary models premised upon exclusive lawyer liability or effective developer immunity inadequately capture these realities and argues for a shared liability framework grounded in foreseeability, control, and the relative capacity to prevent harm as a principled basis for allocating responsibility in AI-assisted legal practice.

RESEARCH QUESTION:

How should liability be allocated between legal professionals and AI developers when generative AI-generated misinformation causes harm within legal practice?

RESEARCH OBJECTIVES:

1. Examine the occurrence and implications of generative AI hallucinations in the context of legal practice.
2. Critically evaluate the extent to which existing principles of professional and developer liability address harms resulting from AI-generated legal misinformation.
3. Analyse comparative approaches to AI accountability to identify emerging trends in the regulation and attribution of responsibility for AI-induced errors.
4. Assess the suitability of a shared liability framework for allocating responsibility between legal professionals and AI developers in instances of AI-generated legal harm.

METHODOLOGY

This study includes a doctrinal and comparative research methodology, examining judicial decisions, professional ethical standards, emerging regulatory frameworks, and contemporary scholarly discourse to assess the legal implications of generative AI hallucinations within legal practice. It proceeds on the premise that existing liability regimes, developed in contemplation of predominantly human error, are ill equipped to address harms occasioned by the probabilistic and opaque nature of generative AI systems. Against this backdrop, the paper advances the argument that a more nuanced allocation of responsibility is required, one that preserves the non-delegable professional obligations of legal practitioners while recognising the corresponding duties of AI developers to ensure transparency, reliability, and adequate risk mitigation in the deployment of legal AI tools.

Literature Review

Academic engagement with generative artificial intelligence in legal practice has largely developed along three interconnected yet insufficiently integrated strands. The first interrogates the phenomenon of AI hallucinations and the epistemic limitations of large language models. Scholars such as Bender, Gebru and their collaborators have highlighted the tendency of large language models to generate linguistically plausible yet factually unreliable outputs, while recent studies examining generative AI in legal contexts have underscored the risks of fabricated authorities, inaccurate citations, and erroneous legal analyses. This body of literature primarily concentrates on issues of model opacity,

verifiability, and the dangers associated with excessive reliance upon predictive systems in professional settings.

A second strand examines the implications of generative AI for legal ethics and professional responsibility. Existing commentary largely converges on the proposition that technological assistance cannot diminish lawyers' established duties of competence, diligence, independent judgment, and candour towards courts and clients. Contemporary professional guidance, particularly ABA Formal Opinion 512 (2024), strengthens this position by emphasizing lawyers' continuing obligation to understand the capabilities and limitations of generative AI and verify AI-assisted outputs independently.

A comparatively emerging body of scholarship considers broader questions of AI governance, developer accountability, and product liability. Emerging studies advocate obligations relating to transparency, safety by design, explainability, and adequate risk disclosure for developers of high risk AI systems. Nevertheless, these discussions frequently analyse professional negligence and developer responsibility as discrete inquiries, offering limited guidance on how liability should be allocated when AI-generated legal misinformation causes harm within legal practice. By interrogating this doctrinal disconnect, the present study seeks to address an important lacuna and assess the viability of a shared liability framework capable of accommodating the distributed risks associated with generative AI in legal services.

SECTION 1

A. Hallucinations as a Distinct Category of Legal Risk

The increasing use of generative artificial intelligence in legal research, drafting, and advisory functions has introduced a novel category of professional risk commonly termed "AI hallucinations." Unlike conventional software errors arising from coding defects or incomplete databases, hallucinations occur when large language models generate factually inaccurate or entirely fabricated outputs based on probabilistic predictions of linguistic patterns rather than verification against authoritative legal sources. Consequently, generative

AI systems may produce non-existent judicial decisions, misquote precedents, inaccurately interpret statutory provisions, or present erroneous legal analyses with persuasive coherence. The legal significance of such errors is amplified by the profession's dependence on verifiable authorities and lawyers' continuing duties of competence, diligence, and candour towards tribunals. As illustrated by subsequent judicial responses to AI-generated fictitious citations, hallucinations challenge traditional assumptions regarding reasonable professional reliance, foreseeability of harm, and the allocation of responsibility between users and developers of generative AI systems.

B. Judicial Encounters including AI Hallucinations

The risks associated with AI hallucinations have rapidly transitioned from theoretical concerns to tangible challenges approach courts and legal practitioners. The seminal decision in **Mata v. Avianca, Inc.** provides the clearest judicial illustration. In that case, counsel submitted a brief containing six fictitious authorities and fabricated quotations generated by ChatGPT, prompting the United States District Court for the Southern District of New York to impose sanctions and reaffirm that attorneys retain a non-delegable duty to verify the accuracy and authenticity of all authorities presented to the court. Comparable concerns emerged in proceedings involving Michael Cohen, where counsel relied upon AI-generated case citations later discovered to be non-existent. Collectively, these incidents demonstrate an emerging judicial tendency to address hallucination-induced harms through established principles of professional responsibility, particularly duties of competence, diligence, and candour towards tribunals. However, while courts have thus far attributed fault primarily to legal practitioners, they have offered limited guidance on whether developers of generative AI systems designed with foreseeable risks of hallucination may also bear responsibility for consequent legal and economic harms.

C. Consequences for Professional Responsibility and the Administration of Justice

AI hallucinations implicate foundational norms of legal practice and adjudication. Fabricated authorities and erroneous legal analyses may prejudice clients, expose practitioners to disciplinary sanctions for breaching duties of competence, diligence, and candour, and impose additional verification burdens upon courts. More broadly, such errors risk undermining public confidence in AI-assisted legal services and the integrity of judicial

proceedings, thereby necessitating a reassessment of conventional approaches to legal accountability.

SECTION 2

A. Traditional Professional Duties as the Basis of Lawyer Liability

Judicial responses to AI-induced legal errors have largely assimilated hallucination-related harms within established doctrines of professional responsibility. Irrespective of the technologies employed, lawyers remain bound by enduring duties of competence, diligence, and candour towards courts and clients. In the United States, these obligations are reflected in Rules 1.1, 1.3, and 3.3 of the ABA Model Rules of Professional Conduct, which require competent representation, reasonable diligence, and truthfulness before tribunals. Comparable duties arise under the Advocates Act, 1961 and the Bar Council of India Rules, which oblige advocates to uphold professional integrity and faithfully protect their clients' interests.

The decision in *Mata v. Avianca, Inc.* judicially affirms this position. There, counsel submitted a brief containing fictitious authorities and fabricated quotations generated by ChatGPT, prompting the United States District Court for the Southern District of New York to impose sanctions and reiterate that attorneys cannot delegate their obligation to verify the authenticity and accuracy of legal authorities relied upon in judicial proceedings. The decision conceptualises generative AI as a sophisticated research aid rather than an autonomous actor capable of displacing traditional principles of legal accountability. Consequently, hallucination-induced harms continue to be addressed primarily through conventional frameworks of professional negligence, notwithstanding the distinctive and foreseeable risks embedded within generative AI systems.

B. Reliance on Generative AI and the Evolving Standard of Care

The increasing integration of generative AI into legal practice necessitates a reassessment of the standard of care expected of reasonably competent practitioners. Traditional negligence principles require lawyers to exercise the degree of skill, prudence, and diligence ordinarily possessed by members of the profession, irrespective of the technological tools employed. Given the well-documented propensity of generative AI systems to hallucinate, uncritical

reliance upon AI-generated legal research or drafting may increasingly fall below this standard. This position is reinforced by ABA Formal Opinion 512 (2024), which emphasizes that lawyers employing generative AI remain responsible for understanding its capabilities and limitations, independently verifying outputs, and exercising appropriate professional judgment. The reasoning adopted in Mata v. Avianca, Inc. similarly suggests that courts are unlikely to regard reliance on AI-generated authorities as a defence to professional misconduct or negligence. Nevertheless, exclusive attribution of fault to legal practitioners risks obscuring the foreseeable and systemic limitations embedded within generative AI systems, thereby inviting closer scrutiny of developer responsibility.

III(C). The Limits of a Lawyer-Centric Liability Model

While existing professional responsibility frameworks appropriately preserve lawyer accountability, an exclusively practitioner-centred approach may inadequately address harms arising from generative AI hallucinations. Courts in cases such as Mata v. Avianca, Inc. have understandably assimilated AI-induced errors within conventional doctrines of professional negligence and disciplinary liability, emphasizing lawyers' continuing duties of competence, diligence, and candour. However, unlike traditional legal research platforms, generative AI systems are intentionally designed, trained, and deployed despite known risks of fabrication and inaccuracy. Consequently, attributing liability solely to practitioners risks overlooking the potential significance of product liability principles, negligent design theories, and duties to warn incumbent upon developers of systems whose limitations are foreseeable and increasingly documented. This doctrinal tension warrants a reassessment of whether existing liability regimes adequately address the distributed nature of harms caused by generative AI.

SECTION 3

A. Theoretical Foundations of Developer Liability

The increasing incorporation of generative artificial intelligence into legal practice necessitates reconsideration of whether liability for hallucinated outputs should remain exclusively with legal professionals or extend, at least proportionately, to AI developers. Developers possess superior knowledge of model architecture, training data, documented failure modes, and available mitigation mechanisms. This informational asymmetry places them in the strongest position to foresee, monitor, and reduce the systemic risks inherent in large language models.

Negligence principles furnish a persuasive doctrinal basis for recognising such responsibility. In ***Donoghue v Stevenson***, the House of Lords established that a duty of care arises where harm is reasonably foreseeable and parties stand in a relationship of sufficient proximity. Likewise, ***MacPherson v Buick Motor Co***. affirmed that manufacturers owe obligations to foreseeable users of products capable of causing injury. By analogy, developers who commercially deploy models despite empirical evidence of their propensity to fabricate authorities, misstate legal propositions, or generate fictitious citations may owe users a duty to exercise reasonable care in the design, testing, deployment, and continuing supervision of such systems. The duty is particularly compelling where developers actively market these tools for professional use while retaining exclusive control over model refinement, safety protocols, and post-deployment updates.

Normative theories further strengthen the case for developer accountability. Corrective justice supports assigning responsibility to actors who substantially contribute to foreseeable risks that subsequently materialise into legally cognisable harm. Enterprise liability similarly requires entities that profit from AI commercialisation to internalise losses arising from systemic defects. From a law-and-economics perspective, developers are also the cheapest cost avoiders, possessing a superior capacity to implement safeguards, enhance reliability, and continuously monitor performance at substantially lower social cost than dispersed end-users. Taken together, these considerations provide a compelling theoretical justification for recognising proportionate developer liability for harms arising from generative AI hallucinations.

B. Product Liability, Negligent Design, and the Duty to Warn

Although generative AI systems do not fit neatly within conventional product liability regimes, established doctrines remain instructive. In Greenman v Yuba Power Products, Inc., the California Supreme Court recognised that manufacturers introducing products into the stream of commerce should bear responsibility for defects causing foreseeable harm. Hallucinations may not constitute traditional manufacturing defects; however, they may plausibly be conceptualised as design defects where developers fail to incorporate reasonable safeguards capable of detecting fabricated legal materials or reducing predictable inaccuracies.

Equally significant is the duty to warn. Boilerplate disclaimers stating that outputs "may be inaccurate" are unlikely to discharge obligations where systems are actively marketed for legal research, drafting, or professional assistance. Developers possessing empirical knowledge regarding known limitations may therefore be expected to provide meaningful disclosures concerning heightened hallucination risks, operational constraints, and the continuing necessity of independent human verification. Emerging risk-based AI governance frameworks further reinforce evolving expectations of transparency, documentation, and post-deployment monitoring.

C. Challenges in Imposing Developer Liability

Notwithstanding these considerations, imposing developer liability presents significant doctrinal difficulties. The opaque and probabilistic nature of large language models complicates the identification of specific design defects and the establishment of causation. Open-source development, intermediary deployment by third parties, contractual limitation clauses, and concerns regarding over-deterrence may further impede liability claims. Consequently, while complete developer immunity appears increasingly inconsistent with contemporary principles of risk allocation, exclusive developer liability would be equally unsatisfactory. Responsibility should instead be apportioned according to control, foreseeability, and the relative capacity of each actor to prevent harm, thereby laying the foundation for a calibrated shared liability framework.

SECTION 4

IV. Towards a Shared Liability Framework for Generative AI Errors in Legal Practice

Neither exclusive lawyer liability nor blanket developer immunity adequately reflects the realities of AI-assisted legal practice. The former disregards developers' superior knowledge of systemic limitations, while the latter permits foreseeable risks to be externalized onto legal professionals and clients. A more coherent allocation of responsibility lies in a framework grounded in foreseeability, control, and the relative capacity to prevent harm, consistent with principles of corrective justice, enterprise liability, and cheapest-cost avoidance.

Lawyers should continue to bear professional and disciplinary responsibility where hallucinated authorities are submitted without reasonable verification, pursuant to ABA Model Rules 1.1, 1.3, and 3.3, ABA Formal Opinion 512, Mata v. Avianca, Inc., and analogous obligations under the Advocates Act, 1961 and the Bar Council of India Rules. Developers, however, should assume proportionate responsibility where foreseeable harms arise from inadequate testing, deficient safeguards, insufficient disclosures, or failures in post-deployment monitoring.

To operationalize this allocation, courts and disciplinary bodies may employ the following inquiry:

FACTOR	GUIDING QUESTION	INDICATIVE LIABILITY
Foreseeability	Was the hallucination a known or reasonably anticipated model limitation?	Developer
Control	Which actor exercised greater control over the relevant risk?	Context dependent
Capacity to prevent harm	Which actor could have prevented the harm at the lowest social cost?	Lawyer, developer or shared

By shifting the inquiry from the mere use of generative artificial intelligence to responsibility for risk creation and mitigation, this framework avoids both the over-deterrent effects of exclusive lawyer liability and the under-deterrent consequences of complete developer immunity. It therefore offers a principled and administrable standard for allocating responsibility in AI-assisted legal practice.

V. **Proposed Framework:** A Three-Factor Test for Allocating Liability for Generative AI Hallucinations

To address the limitations of existing liability regimes, this paper proposes a structured adjudicatory framework for determining responsibility in cases involving hallucinated legal outputs. The framework does not presume fault on the basis of professional status alone. Instead, liability should be apportioned through an assessment of three interrelated considerations derived from negligence principles, enterprise liability, corrective justice, and cheapest-cost-avoider theory.

First, foreseeability: adjudicators should determine whether the hallucinated output arose from a known or reasonably foreseeable limitation of the generative AI system. Evidence of documented hallucination risks, inadequate testing, deficient safeguards, or insufficient disclosures may justify attributing responsibility to developers.

Secondly, control: adjudicators should evaluate which actor exercised greater control over the creation, dissemination, or reliance upon the erroneous output. This inquiry recognises that responsibility may vary depending upon whether the harm principally resulted from deficiencies in system design or a lawyer's failure to independently verify AI-assisted work.

Finally, capacity to prevent harm: adjudicators should assess which actor possessed the superior opportunity to avert the harm at the lowest social cost. This factor accommodates proportionate or shared liability where both lawyers and developers occupied positions enabling effective risk mitigation.

By focusing on responsibility for risk creation, management, and prevention, the proposed framework offers a principled, technologically responsive, and judicially administrable standard for allocating liability in AI-assisted legal practice.

CONCLUSION

Generative artificial intelligence has exposed a fundamental mismatch between traditional liability doctrines and the distributed risks associated with hallucinated outputs in legal practice. Regimes that place exclusive responsibility upon lawyers disregard developers' superior knowledge of model limitations and capacity to anticipate systemic failures, while approaches that effectively immunize developers allow foreseeable harms to be externalized onto practitioners and clients. This paper contends that accountability in AI-assisted legal services should be recalibrated to reflect responsibility for the creation, control, and mitigation of algorithmic risk rather than professional status alone. The proposed Three-Factor Liability Test operationalizes this recalibration by equipping courts and disciplinary bodies with a principled, administrable, and technologically responsive standard for allocating responsibility according to foreseeability, control, and the relative capacity to prevent harm. By aligning liability with risk governance, the framework preserves professional integrity, safeguards clients, and sustains public confidence in the fair administration of justice.

Cases

- Mata v Avianca Inc 678 F Supp 3d 443 (SDNY 2023) law.berkeley.edu/wp-content/uploads/archive/2025/12/Mata-v-Avianca-Inc.pdf

- Donoghue v Stevenson [1932] AC 562 (HL) Donoghue v Stevenson [1932] AC 562 HL: Understanding the Foundation of Negligence Law | Dhyeya Law
- MacPherson v Buick Motor Co. 217 NY 382 (NY 1916) MacPherson v Buick
- Greenman v Yuba Power Products, Inc. 59 Cal 2d 57 (Cal 1963)

Greenman v. Yuba Power Products, Inc. :: :: Supreme Court of California Decisions :: California Case Law :: California Law :: U.S. Law :: Justia

REFRENCES

Legislation and Regulatory Instruments

1. European Parliament and Council Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
2. American Bar Association, Formal Opinion 512: Generative Artificial Intelligence Tools (2024).
3. National Institute of Standards and Technology, AI Risk Management Framework 1.0 (NIST 2023).
4. Ernest J Weinrib, The Idea of Private Law (Harvard University Press 1995)
5. Richard A Posner, Economic Analysis of Law (9th edn, Wolters Kluwer 2014)