UDK 34:004.89

Oleksii SHAMOV
Intelligent systems researcher,
head of Human Rights Educational Guild
18001, Ukraine, Cherkasy, Rizdvyana, 40/28,
yursprava@gmail.com
ORCID: 0009-0009-5001-0526

RESPONSIBILITY FOR ERRORS OF GENERATIVE AI IN LEGAL PRACTICE: ANALYSIS OF "HALLUCINATION" CASES AND PROFESSIONAL ETHICS OF LAWYERS

http://doi.org/10.23939/law2025.48.531

The rapid adoption of generative artificial intelligence (AI) in legal practice has created a significant challenge. While AI tools promise unprecedented efficiency, they are prone to "hallucinations" generating plausible but entirely fabricated information. Recent court cases demonstrate a trend of holding lawyers strictly liable for submitting AI-generated falsehoods, creating an unsustainable professional risk. *Purpose:* This article aims to analyze the current liability framework for errors made by generative AI in legal practice and, based on identified gaps, to propose a new, more balanced model of distributed liability. *Methods:* The research methodology includes a doctrinal analysis of landmark court cases (Mata v. Avianca, Park v. Kim), a systematic analysis of ethical rules and guidance from professional bar associations, and a content analysis of academic publications indexed in Scopus and Web of Science. *Conclusion:* The findings indicate that the current model, which places the entire burden of liability on the lawyer, is untenable. This is due to the empirically proven unreliability of even specialized legal AI tools and the significant legal shields protecting AI developers from liability. The article proposes a novel hypothesis advocating for a shift to a distributed liability model. This model is built on three pillars: (1) a certification system for legal AI tools to guarantee baseline accuracy: (2) a "safe harbor" provision within ethical rules to protect lawyers who use certified tools and follow reasonable verification protocols; and (3) a framework for proportional liability for developers, particularly when their products fail to meet advertised standards. *Further research* should focus on developing specific criteria for AI certification and detailed verification protocols for legal practitioners.

Keywords: generative AI, legal ethics, lawyer liability, AI hallucinations, distributed liability.

Formulation of the problem. Modern legal practice is undergoing a fundamental transformation, catalyzed by artificial intelligence (AI), particularly its generative models. These technologies, capable of creating new content by analyzing vast arrays of data, are permeating all areas of legal activity: from drafting contracts and procedural documents to conducting complex legal research and analyzing evidence. The potential for increased efficiency, reduced costs, and expanded access to legal services is undeniable. However, this technological wave brings not only opportunities but also unprecedented challenges, chief among them being the problem of reliability and liability.

At the core of this challenge lies an inherent property of modern large language models (LLMs), known as "hallucinations." This term, though anthropomorphic, aptly describes the ability of AI to generate information – including citations, references to legal acts, and even entire judicial decisions - that appears entirely plausible but is, in fact, completely fabricated. Unlike traditional databases that retrieve existing information, generative AI predicts the next most likely word in a sequence, which can lead to the creation of syntactically correct but semantically false statements.

This problem has ceased to be purely theoretical and has erupted in courtrooms, creating a series of high-profile precedents. The most resonant case was Mata v. Avianca, Inc. [3], which sent shockwaves through the legal communities of the US and the world. In this case, lawyers representing the plaintiff submitted a brief containing citations and references to six non-existent court cases generated by ChatGPT. When the opposing counsel and the court could not find the cited precedents, the lawyers continued to insist on their existence, relying on further assurances from the chatbot. This case became a stark illustration not only of the technological imperfection of AI but also of a dangerous blind trust in it by professionals. The court's reaction was severe: sanctions were imposed on the lawyers and their firm for violating professional duties, and the judicial opinion emphasized that a lawyer's duty to verify the authenticity of information is absolute and cannot be delegated to a machine.

This and other similar cases have posed a fundamental question to the legal profession: how should liability for errors made not by a human but by an algorithm be allocated in a situation where the use of such algorithms is becoming an economic necessity? The purpose of this study is to conduct a comprehensive analysis of the current liability paradigm for the use of generative AI in legal practice, to identify its systemic shortcomings, and to justify the need for a transition to a new, more balanced model of distributed liability. The objectives of the study are: to conduct an in-depth analysis of key judicial precedents and ethical norms; to study scientific data on the actual reliability of modern AI tools; to investigate the complex legal aspects of the liability of AI development companies; and, as a final result, to formulate and provide a detailed justification for a proposal to implement a distributed liability model that considers the interests of all parties.

An analysis of recent academic publications indicates an active but still fragmented discussion. The work of Stanford researchers [1] provided empirical evidence of the systemic unreliability of even specialized legal AI. Other authors, like Andrew Perlman, explore the evolution of ethical standards, predicting that the duty of technological competence may eventually transform into a duty to use AI [4]. The article "Ethical Lawyering in the Age of Generative AI" [5] argues for the inadequacy of existing ethical rules and the need to develop new ones. However, most existing works either state the problem or analyze the lawyer's liability in isolation, without proposing a holistic, systemic solution. This article aims to fill this gap by proposing a comprehensive model that views the problem not as the fault of an individual lawyer but as a systemic challenge for the entire legal ecosystem.

Analysis of the study of the problem. The contemporary academic and professional discourse on AI in jurisprudence is rich but often focused on specific aspects of the problem, leaving the central dilemma of liability unresolved. This article builds upon key works to outline the contours of the problem and to highlight the unresolved area to which it is dedicated. A foundational work that undermines confidence in AI reliability is the study by Daniel Ho and others from Stanford University, "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools" [1]. This work is critically important because it shifts the discussion from a theoretical to an empirical plane. The authors tested leading legal AI tools designed specifically for lawyers and equipped with Retrieval-Augmented Generation (RAG) technology, which is intended to minimize the risk of "hallucinations" by grounding the model's responses in a specific database of legal documents. The results were alarming: even these advanced systems generated false information in 17-33% of cases. The research methodology involved asking binary (yes/no) legal questions whose answers could be objectively verified. The fact that the error rate is so high even under such controlled conditions indicates a deep, systemic problem. This study proves that the simple advice for lawyers to "use only specialized tools" is insufficient and does not guarantee protection from misinformation.

The geographical dimension of the problem is revealed in the comparative legal analysis "Hallucinations in Legal Practice: A Comparative Case Law Analysis" [2]. This work demonstrates that the challenges associated with AI are not unique to the US. By analyzing practices in the United Kingdom, Canada, Australia, and Pakistan,

the authors show that legal systems worldwide are struggling to find an adequate response. For instance, some Canadian courts have introduced directives requiring lawyers to disclose the use of AI in preparing documents. In the UK, the judiciary has issued guidance emphasizing the lawyer's responsibility but encouraging exploration of the technology. This lack of a unified international standard creates legal uncertainty for global law firms and underscores the need to develop universal principles that could be adapted to national legal systems.

From the perspective of professional ethics, works analyzing the Model Rules of Professional Conduct of the American Bar Association (ABA) are central. The article "Ethical Lawyering in the Age of Generative AI" [5] is illustrative in this regard. The author argues that while existing rules, such as the duty of competence (Rule 1.1) and candor toward the tribunal (Rule 3.3), are formally applicable, they were not designed with the unique risks of AI in mind, such as the scalability of errors and the opacity of algorithms. The main argument is that a new, specialized rule is needed to clearly regulate the use of AI, similar to how special rules exist for handling client trust accounts.

However, the unresolved part of the overall problem, to which this article is dedicated, remains the systemic link between the lawyer's liability and the developer's liability. On the one hand, there are deep analytical reports, such as the RAND Corporation's "Liability for Harms from AI Systems" [6], which detail potential scenarios for holding developers liable under tort law. They analyze complex issues like the standard for proving negligence and the possibility of applying strict liability. On the other hand, there are numerous practical guides for lawyers, such as the one published by David Levine [10], which focus on internal risk management procedures, like developing firm policies and staff training.

The gap lies in the absence of a model that connects these two worlds. It is not enough to simply say that lawyers must be careful and developers could potentially be held liable. A functional system is needed that defines how these two types of liability relate, how they can mutually balance each other, and how to create incentives for both parties to minimize risks. This article aims to fill this gap by developing a comprehensive concept of distributed liability.

The purpose of the article. To achieve the stated purpose and ensure a comprehensive analysis of the problem, a complex methodology was employed, combining several complementary scientific methods. The choice of methods was dictated by the interdisciplinary nature of the research, which lies at the intersection of law, technology, and professional ethics.

The foundation of the research was doctrinal analysis, which is traditional for legal studies. This method was applied for a deep study and interpretation of primary legal sources. The objects of analysis were the court decisions in Mata v. Avianca, Inc. [3] and Park v. Kim [8], which are cornerstones in shaping judicial practice regarding AI. The analysis covered not only the final conclusions of the courts regarding sanctions but also the reasoning in the opinions, where judges justified the application of Rule 11 of the Federal Rules of Civil Procedure and clarified the inviolability of the lawyer's duty to conduct a "reasonable inquiry." This method allowed for an understanding of the legal logic that places the burden of liability on the lawyer.

Systematic analysis was used to study the regulatory landscape created by professional legal associations. Instead of viewing each ethical rule in isolation, this method allowed for them to be seen as a single system. The ABA Model Rules of Professional Conduct and the official opinions and guidelines issued by the bar associations of California, Florida, and New York were analyzed. The interrelation between the duties of competence (Rule 1.1), confidentiality (Rule 1.6), candor toward the tribunal (Rule 3.3), and supervision (Rules 5.1 and 5.3) was examined in the context of AI use. This helped to understand how the existing ethical infrastructure is adapting to new technological challenges.

An important component was the content analysis of academic and expert publications. Eleven key sources were selected that met the criteria of relevance (published mainly in the last 3-5 years), authoritativeness, and scientific value. Priority was given to publications indexed in international scientometric databases such as Scopus and Web of Science [1, 4, 5], as well as reports from leading think tanks (RAND Corporation, Stanford HAI) [6, 7]. The content analysis was aimed at identifying key arguments, empirical data, theoretical concepts, and legal proposals that exist in the current discourse.

To understand the global context of the problem, elements of the comparative-legal method were applied, allowing the research to extend beyond the US legal system. The analysis of an article examining practices in different countries [2] helped to identify common problems and differences in regulatory approaches, which confirmed the universality of the challenge.

In the final stage, the method of synthesis was used. The results obtained through the aforementioned methods were combined to form a holistic picture of the problem. It was at the synthesis stage that the key contradiction was identified between the absolute liability of the lawyer and the systemic unreliability of the technology, combined with the limited liability of the developer. This synthesis became the basis for developing the scientific hypothesis about the need to transition to a distributed liability model, which serves as the main scientific novelty of this article.

Presenting main material. The analysis of the collected materials reveals a deep and unstable asymmetry in the distribution of liability for the errors of generative AI. This asymmetry creates a paradoxical situation where lawyers are held almost absolutely responsible for the results of tools they do not control and which are demonstrably unreliable, while the creators of these tools are largely shielded from liability.

Part 1. The Status Quo: The Lawyer as the Sole Center of Liability. Judicial practice and ethical norms have formed a clear and rigid approach: the end-user, i.e., the lawyer, bears full responsibility for the accuracy of information submitted to a court. The case of Mata v. Avianca [3] became definitive in this respect. In his decision, the judge not only imposed financial sanctions but also conducted a detailed analysis of the lawyers' conduct, highlighting several key violations. First, the basic duty of candor toward the tribunal was violated, as the lawyers continued to defend the existence of fictitious cases even after they were questioned. Second, the judge pointed to a gross violation of Rule 11 of the Federal Rules of Civil Procedure, which requires that every document filed with the court be supported by an "inquiry reasonable under the circumstances." The court explicitly stated that simply asking a chatbot "is this a real case?" does not constitute a reasonable inquiry.

The case of Park v. Kim [8], decided by an appellate court, only strengthened this approach. In this case, the court clearly stated that when an attorney cites a precedent, they are personally certifying to the court that they have read that precedent and believe it to be relevant. The attempt to shift the blame to ChatGPT was dismissed as baseless. These cases create a powerful precedent that effectively establishes a strict liability regime for the lawyer: it does not matter how the error occurred - its presence in a court document is a violation of professional duty.

Ethics committees of leading bar associations support this position. In their guidance, they clarify that the existing ABA Model Rules are fully applicable. The duty of competence (Rule 1.1) requires not only knowledge of the law but also an understanding of the benefits and risks of associated technologies. This means a lawyer cannot plead "ignorance" of how AI works.

The duty of supervision (Rules 5.1 and 5.3), as noted by David Levine in their analysis [10], extends not only to junior lawyers but also to "non-lawyer assistants," which can certainly include AI platforms. This means a senior partner is responsible for ensuring that their firm uses reliable processes for verifying the outputs of AI work.

Part 2. The Technological Paradox and the "Legal Shield" of Developers.

This rigid liability regime for lawyers comes into glaring conflict with two factors: the proven unreliability of the technology and the limited liability of its developers. As the Stanford University study showed [1], even the best legal AI platforms, which cost significant money and are advertised as reliable, systematically make mistakes. This is due to the very nature of large language models, which are probabilistic, not deterministic, systems.

Demanding that a lawyer guarantee 100 % accuracy of a result obtained from such a tool is like demanding that a meteorologist guarantee a perfectly accurate weather forecast a month in advance.

Meanwhile, holding development companies like OpenAI, Google, or Microsoft liable is extremely difficult. The RAND Corporation article [6] and the analysis from Stanford HAI [7] highlight several layers of legal protection for developers.

· The "Product vs. Service" Dilemma. In US tort law, there is a significant difference between liability for a defective product and for a negligently provided service. For products, a strict liability regime often applies, where the plaintiff does not need to prove the manufacturer's fault, only the existence of a

defect and causation. For services, a negligence standard applies, requiring proof that the service provider acted below a reasonable professional standard. It is still unclear how courts will classify generative AT, which creates a significant hurdle for plaintiffs.

· Section 230 Immunity. Section 230 of the Communications Decency Act protects online platforms from liability for content created by third parties. AI developers may argue that they are merely platform operators and that the content is generated by the model itself (or in response to a user's prompt), thus shielding them with immunity. Although the applicability of this immunity to generative AI is a subject of heated debate, it remains a powerful potential defense.

· First Amendment Protection. Some experts suggest that developers might argue that the output of their models is a form of "speech" or "expression" and therefore falls under the protection of the First Amendment to the U.S. Constitution, which guarantees freedom of speech.

These factors create a "liability vacuum" in which the lawyer, at the end of the chain, becomes the only party to whom sanctions can be easily applied.

Part 3. The Proposal: A Distributed Liability Model as the Way Forward. To escape this unstable situation, it is necessary to fundamentally rethink the approach to liability. Instead of seeking a single culprit, a system should be created that distributes responsibility among all participants in the process and incentivizes them toward safe behavior. This article proposes a model based on three pillars.

1. Creation of a Certification and Standardization System for Legal AI.

Analogous to how standards exist for medical equipment or financial instruments, a certification system for legal AI platforms should be developed. This process could be led by an independent regulator or a specialized body within a professional association (e.g., the ABA or Ukrainian National Bar Association). To obtain certification, a developer would have to undergo an independent audit that verifies its product against a range of criteria: citation accuracy, hallucination rates on standardized tests, reliability of confidential client data protection, and transparency regarding the sources on which the model was trained. The audit results and the confirmed error rate should be public. This would allow lawyers to make an informed choice and developers to compete not just on functionality but also on reliability.

2. Introduction of a "Safe Harbor" for Lawyers. This element is key to encouraging innovation. If a lawyer uses a certified AI tool for its intended purpose and adheres to a clearly defined "protocol of reasonable verification", they should be protected from sanctions if an error still occurs. The concept of a "safe harbor", as noted in the analysis of the DMCA rules [9], does not completely absolve one of liability but creates conditions under which good-faith conduct is protected. A verification protocol might include these steps: a) mandatory checking of all citations to judicial decisions using a traditional, reliable database (e.g., Westlaw, LexisNexis); b) critical evaluation of the logical structure and soundness of legal arguments generated by AI; c) cross-checking of key factual data (names, dates, amounts) against primary sources. This would replace the vague requirement to "check everything" with a concrete and feasible set of actions.

3. Establishment of Proportional Liability for Developers. This pillar balances the system. A developer's liability should depend on their actions and claims. If a developer markets an uncertified tool while misleading users about its reliability, principles of liability for misrepresentation or negligence could apply. If a certified tool systematically performs worse than claimed during certification, this could be grounds for revoking the certificate and for lawsuits seeking damages. Academic proposals for creating special liability regimes, such as mandatory insurance or collective funds to cover risks, could be integrated into this model. This would create a direct financial incentive for developers to invest in improving the quality and safety of their products, not just in marketing.

Conclusions. The conducted research leads to the conclusion that the current approach to liability for the errors of generative AI in legal practice is reactive, asymmetrical, and unsustainable. It places the entire burden on lawyers, forcing them to be responsible for the flaws of technologies they did not create and do not control. This situation not only creates unfair pressure on the legal profession but also hinders the process of healthy innovation, forcing lawyers either to completely abandon potentially useful tools or to use them with excessive risk.

In contrast to this approach, the article argues for the necessity of transitioning to a proactive and balanced model of distributed liability. This model recognizes that ensuring reliability and safety in the age of AI is a shared task for all participants in the legal ecosystem. The proposed three-component structure, which includes the certification of AI tools, the creation of a "safe harbor" for conscientious lawyers, and the introduction of proportional liability for developers, is designed to transform the current paradigm. It shifts the focus from punishment for errors to creating conditions for their prevention.

The implementation of such a model would have far-reaching positive consequences. For lawyers, it would create clear and feasible rules of the game, allowing them to leverage the benefits of AI with manageable risk. For developers, it would create strong market incentives to invest in the quality, reliability, and transparency of their products. For the legal system as a whole, it would promote increased trust in technology and ensure its more harmonious and safe integration into the administration of justice.

Prospects for further research in this area are broad and multifaceted. First, there is an urgent need to develop a concrete technical-legal framework for AI certification. This will require interdisciplinary collaboration among lawyers, IT engineers, ethicists, and representatives of regulatory bodies to define measurable criteria for accuracy, reliability, and security. Second, the content of the reasonable verification protocols that would form the basis of the "safe harbor" needs to be detailed and formalized, possibly through amendments to professional ethics codes. Third, an important direction is the analysis of the economic aspects of the proposed model, including the development of liability insurance models for both lawyers and developers.

Finally, it is critically important to research ways to adapt and implement this model in different national legal systems, particularly in Ukraine, to ensure its compliance with local legislation and the needs of the legal community.

Acknowledgements None.

Funding The author declares no financial support for the research,authorship, or publication of this article.

Author contributions The author confirms sole responsibility for this work. The author approves of this work and takes responsibility for its integrity.

Conflict of interest The author declares no conflict of interest

REFERENCES

1. Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. & Ho, D. (2025). ***Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools.*** *Journal of Empirical Legal Studies.* № 22. P.1–27. https://doi.org/10.1111/jels.12413 [In English].

2. Munir, B. (2025). Hallucinations in Legal Practice: A Comparative Case Law Analysis. International Journal of Law, Ethics, and Technology. http://dx.doi.org/10.2139/ssrn.5265375 [In English]

3. Justia U.S. Law. (2023). Mata v. Avianca Case. Justia U.S. Law Website. https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/ [In English].

4. Perlman, A. (2024). The Legal Ethics of Generative AI. Suffolk University Law School Research Paper 24-17. http://dx.doi.org/10.2139/ssrn.4735389 [In English].

5. Chang, M. (2025). Ethical Lawyering in the Age of Generative AI. Seattle Journal of Technology, Environmental, & Innovation Law. Vol. 15: Iss. 2, Article 4. https://digitalcommons.law.seattleu.edu/sjteil/vol15/iss2/4 [In English].

6. Smith, G., Stanley, K., Marcinek, K., Cormarie, P. & Gunashekar, S. (2024). Liability for Harms from AI Systems. RAND Corporation Website. https://www.rand.org/pubs/research_reports/RRA3243-4.html [In English].

7. Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2023). Who Is Liable When Generative AI Says Something Harmful? Stanford University Website. https://hai.stanford.edu/news/ who-liable-when-generative-ai-says-something-harmful [In English].

8. Justia U.S. Law. (2023). Park v. Kim Case. Justia U.S. Law Website. https://law.justia.com/cases/federal/appellate-courts/ca2/22-2057/22-2057-2024-01-30.html [In English].

9. Hickey, K. (2020). Digital Millennium Copyright Act (DMCA) Safe Harbor Provisions for Online Service Providers: A Legal Overview. Library of Congress Website. https://www.congress.gov/crs-product/IF11478 [In English].

10. Levine, D. (2025) Avoiding Ethical Pitfalls as Generative Artificial Intelligence Transforms the Practice of Litigation. The National Law Review. https://natlawreview.com/article/avoiding-ethical-pitfalls-generative-artificial-intelligence-transforms-practice [In English].

Received: 19.07.2025
Revised: 10.09.2025
Accepted: 29.09.2025
Published (online):12.12.2025
Printed:26.12.2025

Олексій ШАМОВ
Дослідник інтелектуальних систем,
Голова ГО “Просвітня фундація з прав людини”
18001, Україна, Черкаси, Різдвяна, 40/28,
yursprava@gmail.com
ORCID: 0009-0009-5001-0526

ВІДПОВІДАЛЬНІСТЬ ЗА ПОМИЛКИ ГЕНЕРАТИВНОГО ШІ В ЮРИДИЧНІЙ ПРАКТИЦІ: АНАЛІЗ СПРАВ ПРО «ГАЛЮЦИНАЦІЇ» ТА ПРОФЕСІЙНА ЕТИКА ЮРИСТІВ

Актуальність стрімкої інтеграції інструментів генеративного штучного інтелекту (ШІ) в юридичну практику створює парадокс: з одного боку, ШІ пропонує безпрецедентну ефективність у роботі з текстами та аналізі даних, з іншого - породжує нові етичні та правові ризики, зокрема через феномен «галюцинацій» - генерування неправдивої інформації.

Судова практика почала формувати підходи до відповідальності за такі помилки, покладаючи весь тягар на юристів, що створює нестійку та ризиковану ситуацію для професії. *Мета* статті - проаналізувати поточну модель відповідальності юристів за помилки генеративного ШІ та, на основі виявлених прогалин, запропонувати нову, більш збалансовану концепцію розподіленої відповідальності. *Методи* дослідження включають доктринальний аналіз судових прецедентів (зокрема, справ Mata v. Avianca та Park v. Kim), системний аналіз етичних норм та керівних принципів професійних юридичних асоціацій, а також аналіз змісту наукових публікацій, що індексуються в базах Scopus та Web of Science. У *висновку*, дослідження показало, що існуюча модель відповідальності, яка покладає її виключно на юриста, є нежиттєздатною через доведену ненадійність навіть спеціалізованих юридичних ШІ-інструментів та значні правові перешкоди для притягнення до відповідальності їх розробників. Запропоновано гіпотезу про необхідність переходу до моделі розподіленої відповідальності, що базується на трьох елементах: сертифікації ШІ-інструментів, запровадженні «безпечної гавані» для юристів та встановленні пропорційної відповідальності для розробників. *Перспективи* подальших досліджень полягають у розробці конкретних критеріїв для сертифікації юридичних ШІ-платформ та детальних протоколів верифікації для юристів, що дозволить практично реалізувати запропоновану модель.

Ключові слова: **генеративний ШІ, юридична етика, відповідальність юриста, галюцинації ШІ, розподілена відповідальність.**