HALLUCINATIONS IN LEGAL PRACTICE: A COMPARATIVE CASE LAW ANALYSIS

Abstract

The employment of Artificial Intelligence (AI) in legal operations raised concerns about ethical challenges and their potential consequences. Among other issues, hallucinations refer to a phenomenon whereby AI systems generate plausible but inaccurate or fabricated responses. In legal matters, where precision and compliance with authorities are paramount, inconsistency with legal doctrines and judicial precedents may lead to wrong legal advice or decisions. AI tools such as ChatGPT and Lexis +AI exhibit human-like intelligence. Still, their fabricated responses could lead to real-world consequences such as professional misconduct resulting in civil liabilities. This article contributes to the following aspects: it compares judicial scholarship evolved on AI hallucinations in the USA, Pakistan, UK, Australia, and Canada. It examines the standing orders and policy guidelines set by the bar and bench constituting patchwork with competing outcomes. The article emphasizes uniform and comprehensive policy guidelines for the responsible use of generative AI tools in legal operations.

Keywords

Cases of Hallucination, Standing Orders on Hallucinations, Generative AI, AI and Malpractices, AI and Civil Liabilities

1. Introduction

The recent integration of AI into legal operations offers unparalleled opportunities and poses critical challenges.[1] Though AI models are instrumental in performing legal tasks, their adoption is hampered by crucial concerns such as producing incorrect or deceptive outcomes, commonly known as hallucinations.[2] Modern AI solutions are transforming the legal fields, including legal education, research, and practice.[3] Within a few months of its public release in November 2022, ChatGPT secured itself as the fastest-ever growing consumer application in human history.[4] Embracing the trend, recent studies have found that generative AI witnessed remarkable performance in law school exams, Bar exams, and other legal analyses.[5] AI enables machines to mimic human intelligence, empowering them to learn, solve problems, and make decisions. The employment of AI in various spheres is driving transformative changes and has the potential to revolutionize legal operations. Lawyers are utilizing AI in legal operations to augment legal services. 41 of the top 100 US law firms have initiated AI in their legal services.[6] According to a study by LexisNexis, 80% of Fortune 1000 executives desire their external counsels to enhance efficiency by leveraging AI capabilities. However, these tools are not risk-free and constitute ethical challenges such as bias, copyright, data invasion, fabricated responses, and information security, posing ultimate liability to corroborate their outcomes.[7]

---

[1] Darla Wynon Kite-Jackson, *2023 Artificial Intelligence (AI) TechReport*, AM. BAR ASS'N (Jan 15, 2024).

[2] Matthew Dahl et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, 16 J. LEGAL ANALYSIS 64, (2024), https://academic.oup.com/jla/article/16/1/64/7699227.

[3] Jonathan H. Choi and Daniel Schwarcz, 2024. *AI Assistance in Legal Analysis: An Empirical Study*. J. LEGAL EDUC. doi: 10.2139/ssrn.4539836. (forthcoming), https://elsevier-ssrn-document-store-prod.s3.amazonaws.com/23/08/13/ssrn_id4539836_code499486.pdf; *See also*, Michael A. Livermore, Felix Herron, & Daniel Rockmore, *Language Model Interpretability and Empirical Legal Studies*. J. INSTITUT. THEORETI. ECON., forthcoming (2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4599212; *See also*, Ian Rodgers, John Armour, & Mari Sako, *How Technology Is (or Is Not) Transforming Law Firms, 19* ANN. R. LAW SOCIAL SCI. *299–317* (2023), https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-111522-074716.

[4] See Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, REUTERS (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

[5] Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan, & Daniel Schwarcz, *ChatGPT Goes to Law School*, 71 J. LEGAL ED. 387 (2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4335905; *See also,* Chung Kwan, *What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature*, 13 EDUC. SCI. 410 (2023); John Ney et al., *Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence, Philosophical Transactions A 382(2270),* THE ROYAL SOCIETY, (Feb. 26, 2024), https://www.researchgate.net/publication/378489936_Large_language_models_as_tax_attorneys_a_case_study_in_legal_capabilities_emergence.

[6] Justin Henry, *We Asked Every Am Law 100 Law Firm How They're Using Gen AI. Here's What We Learned*, AM. LAW. (Jan. 29, 2024), https://www.law.com/americanlawyer/2024/01/29/we-asked-every-am-law-100-firm-how-theyre-using-gen-ai-heres-what-we-learned/?slreturn=20241013185149.

[7] Joseph J. Avery, Patricia Sánchez Abril & Alissa del Riego, *ChatGPT, Esq.: Recasting Unauthorized Practice of Law in the Era of Generative AI*, 26 YALE J. L. & TECH. 64 (2023), https://yjolt.org/sites/default/files/avery_abril_delriego_26yalejltech64.pdf; *see also, Amy B. Cyphert, A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law, 55 UC DAVIS L. REV. 401 (2021), https://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/55-1_Cyphert.pdf; Ed Walters, The Model Rules of Autonomous Conduct: Ethical Responsibilities of Lawyers and Artificial Intelligence, 35 Ga. St. U. L. Rev. 1073 (2019), https://readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=2974&context=gsulr; Nicole Yamane, Artificial Intelligence in the Legal Field and the Indispensable Human Element Legal Ethics Demands,* 33 GEO.

Hallucination refers to false or deceptive outputs that AI models perpetuate for various reasons such as insufficient data training, incorrect assumptions, or biases in the dataset. AI models are trained on data and learn to make predictions by finding patterns in the data. The precision of outcomes is often subject to the quality and completeness of the training data. Where the training data is incomplete, biased, or otherwise flawed, the AI models may learn incorrect patterns, leading to inaccurate predictions or plausible fabricating links to webpages that never existed. While considering the efficiencies of AI solutions, new ethical challenges have been posed.[8]

1.1. Generative AI and its Tendency Towards Hallucinations

Given its functions, generative AI is a particular kind of AI that focuses on producing original content in response to users' questions. Generative AI is based on machine learning models, also known as deep learning models, which are algorithms that mimic the human brain's learning and decision-making process. These models learn patterns and structures from the training data and utilize them to comprehend users' natural language prompts and respond with new relevant content. The use of Generative AI became more active with the development of Large Language Models (LLMs), which can generate human-like text based on the features learned from the huge data on which these models are trained. By predicting the next element in a sequence, these models produce new content and host inherent challenges such as perpetuating misinformation.[9] Generative AI may produce erroneous output based on its probabilistic algorithms for making inferences. These models perpetuate the most probable response to a user's prompt without guaranteeing correctness, which may lead to a plausible but fabricated outcome.[10]

LLMs are advanced AI systems that fall under Natural Language Processing (NLP) and are designed to comprehend and produce human language. These models are trained on huge data to learn the intricacies of language by employing transformer architectures, which have revolutionized NLP and other AI tasks since their inception in 2017.[11] These models excel in tasks such as summarizing text, answering questions, and engaging in conversations by generating relevant and coherent text based on their input. For instance, ChatGPT-4 is an LLM developed by OpenAI. However, other generative AI tools such as Microsoft Copilot, Lexis +AI, and Westlaw Co-Counsel leverage the capabilities of LLMs to perform multiple tasks but are not LLMs themselves.

1.2. Modes of Legal Hallucinations

J. LEGAL ETHICS 877 (2020), https://www.law.georgetown.edu/legal-ethics-journal/wp-content/uploads/sites/24/2020/09/GT-GJLE200038.pdf.

[8] Frances Green & Rebecca Porter, *The Legal Vision for the Future or an AI Hallucination? Navigating the Complexities of Attorney Ethics and Use of Artificial Intelligence*, NEW YORK L. J., (April 2, 2024), https://www.law.com/newyorklawjournal/2024/04/02/the-legal-vision-for-the-future-or-an-ai-hallucination-navigating-the-complexities-of-attorney-ethics-and-use-of-artificial-intelligence/?slreturn=20241010143515.

[9] IBM, Generative AI, (last visited Dec. 16, 2024), https://www.ibm.com/topics/generative-ai.

[10] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch & Patrick Zschech, Generative AI, 66 BUS. & INFO. SYS. ENG'G 111 (2024), https://link.springer.com/article/10.1007/s12599-023-00834-7.

[11] Ashish Vaswani et al., *Attention is All You Need*, 30 Advances in Neural Info. Processing Sys. 5998 (2017), https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Attention+is+All+You+Need%22+by+Vaswani+et+al.+in+2017%2C&btnG=.

Undoubtedly, AI systems have the potential to address complex legal tasks but are limited by a notable issue: their tendency to produce incorrect or misleading outcomes.[12] Legal hallucinations can be referred to as the phenomenon where LLMs perpetuate fabricated legal responses, which could be problematic in the legal context where accuracy is paramount. Legal hallucinations are exhibited in many ways such as inventing fictitious precedents, nonexistent statutes, misinterpreting laws, offering inaccurate legal advice, and producing made-up legal content, which hosts various risks, including legal liability, malpractice, and miscarriage of justice.

Legal professionals are increasingly getting involved with AI chatbots without fully realizing how they work and their susceptibility to errors. Even if legal professionals are unwilling to deploy AI, they still need to learn and live with them. Legal professionals are expected to act as custodians of the legal system and should be capable of identifying errors in the outcomes of these models.[13]

Hallucinations occur when AI systems produce incorrect, misleading, or entirely fabricated content: Incorrect predictions, to predict the happening of an unlikely event such as the rain forecast when it does not rain. False positive, to identify something as a threat when it is not such as detecting a fraudulent activity when it is not. False negative, fails to identify something as a threat when it is a threat such as failing to identify a cancerous tumor. Hallucinations could be in any of the following forms: (1) Factual hallucinations, AI systems might produce information factually incorrect or nonexistent such as discovering scientific facts or historical events that are not true.[14] (2) Contextual hallucinations, where AI models misunderstand the context or misinterpret the user's intent. It comes to the fore where AI responses are contextually irrelevant or inappropriate to the given prompt.[15] (3) Logical Hallucinations, where AI responses are logically inconsistent or contradictory. For example, where AI-generated content lacks a coherent line of reasoning. (4) Visual hallucinations, where AI systems generate images containing elements other than input data or distorted unrealistically.[16] (5) Conversational hallucinations, where the AI system fabricates part of a conversation like contributing statements to the people or inventing quotes who never made them.

Like other fields, the recent adoption of LLMs into legal operations offers significant opportunities and considerable challenges.[17]

2. Why Do AI Models Hallucinate?

---

[12] Matthew Dahl, et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models,* arXiv:2401.01301v2 [cs.CL], (Jun 21, 2024), https://arxiv.org/abs/2401.01301.

[13] David Rubenstein, *2024 Selected Topics and Miscellany CLE,* Washburn University School of Law, Presentation *(June 13, 2024), https://www.washburnlaw.edu/about/community/cle/_files/selected-topics-schedule.pdf.*

[14] Ankit, *What is AI Hallucination? Understanding and Mitigating AI Hallucination*, GeeksforGeeks (Jan. 27, 2025), https://www.geeksforgeeks.org/what-is-ai-hallucination/

[15] MIT Sloan Educational Technology Office, *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias.

[16] IBM, *What are AI Hallucinations?,* https://www.ibm.com/think/topics/ai-hallucinations.

[17] Darla Wynon Kite-Jackson, *2023 Artificial Intelligence (AI) TechReport*, ABA TECHREPORT 2023, (Jan. 15, 2024), https://www.americanbar.org/groups/law_practice/resources/tech-report/2023/2023-artificial-intelligence-ai-techreport/.

With the widespread proliferation of AI systems, some critical challenges such as hallucinations have been confronted. Hallucinations result when AI models generate content that is not grounded or realistic. Consequently, AI models might fabricate responses that do not correspond to real-world data, potentially leading to dire consequences. Conventionally, AI hallucinations transpire the way AI models are trained. Most LLMs depend on the data available on the internet, which might contain both correct and incorrect content supplemented with inherent cultural and societal biases. The models mimic patterns from that data without recognizing their truthfulness and can perpetuate imprecision or biases.[18]

From the above conception, intriguing questions arise: Why do we expect AI to be 100% unbiased when humans themselves are not? Why is the burden of absolute accuracy placed on AI programs? It is worth considering why we hold AI to such high standards when, in human-to-human interactions, achieving complete impartiality and accuracy is impossible.

The LLMs are subject to limitations and work like advanced autocomplete tools – designed to foresee the next sequence or word based on the observed patterns – with the underlying objective of creating credible content and not verifying its truthiness. Inversely, any accuracy in their generated content is often inadvertent and might produce output that looks plausible but could be erroneous.[19]

As LLMs by design cannot distinguish between true and false even if these models are trained exclusively on accurate data, there is still a probability of producing new, potentially erroneous content by assimilation of patterns in an unexpected manner.[20] These LLMs are not infallible, and their most puzzling behavior is the production of hallucinations, either incorrect responses or entirely fabricated results that could create real-world challenges where accuracy is paramount. Considering these algorithms are not sentient and cannot distinguish between truth and lies, it is imperative to comprehend the nature of these hallucinations to harness the effectiveness and efficiency of these models responsibly. Though these models might appear sentient because they generate coherent and relevant text, it is notable that they cannot differentiate between truth and false, nor have intentions or beliefs. So, hallucinations are the byproduct of the models' probabilistic nature and limitations in the training data, rather than a deliberate act.[21]

Likewise, a deliberate act of the designer can cause the models to perpetuate inaccurate responses. For instance, data poisoning is an intentional cyberattack, which can degrade the model's performance or cause it to produce incorrect or biased outcomes. Data poisoning can

---

[18] Karen Weise & Cade Metz, *When A.I. Chatbots Hallucinate*, THE NEW YORK TIMES, (May 1, 2023), https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.

[19] Matt O'Brien, *Chatbots Sometimes Make Things Up. Is AI's Hallucination Problem Fixable?*, AP NEWS, (August 1, 2023), https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4.

[20] *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, MIT SLOAN TEACHING & LEARNING TECHNOLOGIES, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias.

[21] Daniel A. Tysver, *AI Hallucinations (Why would I lie?)*, BITLAW, https://www.bitlaw.com/ai/hallucinations-and-AI.html.

occur by modifying existing data by injecting misleading information into the training dataset or deleting a portion to skew results.[22] These attacks aim to manipulate specific outcomes or to degrade the overall robustness of the model's performance.[23] To overcome the issue of data poisoning, it is critical to maintain the quality and integrity of the training data and employ robust security measures.

3. Legal Hallucinations: A Comparative Case Law Study

The following segment provides a detailed analysis of the judicial scholarship that evolved on legal hallucinations and its potential impacts:

3.1. Case Law Development in the USA

New York attorneys faced legal consequences for presenting a brief with fictitious case law precedents generated through ChatGPT.[24] Two lawyers were each sanctioned to pay a $ 5,000 fine for providing a legal brief that referred to six fictitious case citations produced by ChatGPT, which the court regarded to have acted in bad faith by declaring it as an act of conscious avoidance and false and misleading statements to the court. The lawyers used ChatGPT to prepare a personal injury case against Columbian airline Avianca and included references of false citations. The court dismissed the claim on the pretext of statutory limitation. While imposing the sanction, the court declared that using AI is not inherently improper, but the ethics rule requires the attorneys to ensure accuracy in their filings. The lawyers kept standing by their fake opinions despite the court and the airline having questioned the existence of the citations.[25]

Shortly after the New York case, *Ex parte Lee,* another fabricated case, was reported in a Texas appellate court.[26] Allen Michael Lee was charged with three sexual assaults for which the bail was set at $ 400,000, which Lee contested by filing a pre-trial application for the writ of habeas corpus for either his release or reduction of bail to $ 15,000, which the trial court refused. Hence, he challenged the court order at the Court of Appeals of Texas. The court denied review based on the deficient briefing, citing five cases, three of which did not exist in the Southwest Reporter. The court realized the cited cases did not exist and the two others were from the Missouri court, making them immaterial to the argument in the brief. Lee, however, did not address those issues through a reply or a supplemented brief. The court called the brief illogical and at least partly prepared with the help of AI.[27] Unlike the New York case, the court in the instant case did not

---

[22] Bart Lenaerts-Bergmans, *Data Poisoning: The Exploitation of Generative AI*, CROWDSTRIKE, Mar. 20, 2024, https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/data-poisoning/.

[23] Tom Krantz, *What is Data Poisoning?*, IBM, https://www.ibm.com/think/topics/data-poisoning.

[24] Benjamin Weiser, *Here's What Happens When Your Lawyer Uses ChatGPT*, THE NEW YORK TIMES, (May 27, 2023), https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html. *See for details*, Mata v. Avianca, Inc., No. 1:2022cv01461, Document 55 (S.D.N.Y. 2023).

[25] Hon. John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawyers Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40 GA. ST. U. L. REV. 917, 925(2024), https://readingroom.law.gsu.edu/gsulr/vol40/iss4/9/; See also, Sara Merken, *New York lawyers sanctioned for using fake ChatGPT cases in legal brief*, Reuters (June, 26, 2023), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/.

[26] Ex parte Lee, 673 S.W.3d 755, 756 (Tex. App. 2023).

[27] Ibid.

issue a show cause order and report the authority for disciplinary action to the State Bar of Texas because it had addressed the issue raised in the appeal.

In *People v. Crabill*[28], Zachariah Crabill, a young attorney, filed a brief supported by dozens of cases prepared with the assistance of ChatGPT. On the hearing day, he realized the cases he submitted were not available on the LexisNexis and were 'garbage'. He compounded his mistake by not validating the citations or alerting the court and withdrawing the motion, Crabill blamed an intern when the court pointed out the made-up citations. While rejecting the motion, the court referred him to disciplinary action. After six days, Crabill filed an affidavit confessing the use of ChatGPT while drafting the motion. For his professional misconduct, he was terminated from his law firm and banned for one year and one day from practicing law.[29]

In April 2023, Lydia Nicholsen, a Los Angeles housing attorney, realized that the brief in an eviction case received from the opposing counsel, Dennis Block, was supported by fabricated citations. Nicholsen filed a motion and pointed out the fake cases. On confirmation, the judge declared the filings "rife with inaccurate and false statements" and imposed a fine of $ 999 on the firm, which was just under the threshold required for reporting to the state bar for further investigation.[30] The firm shifted liability onto a first-year lawyer, who had since left the firm, by blaming him for relying on an online search.[31]

In *United States v. Michel Cohen*[32], the defense attorney used AI while filing a motion for early release. Cohen, a former attorney for President Donald Trump, confessed to hush money to two women during the presidential campaign. Since November 2021, Cohen has been on supervised release after serving time in prison. His lawyer, Schwartz, filed a motion for his early release. Afterward, another attorney, Danya Perry, was added to the Choen's legal team who realized fabricated citations and alerted the court accordingly. The court issued a show cause notice to Schwartz to provide copies of the three cited cases or respond to why he should not be sanctioned. Based on attorney-client privilege, Schwartz requested to file a response under seal, which was unsealed on December 29, 2023. It was disclosed through a sworn declaration of Cohen that the citations were produced by Google Bard, which Schwartz incorporated with his submission without verification. Cohen, who was disbarred from practice years ago, admitted that he had not kept up with the trends in legal technology of these tools to produce citations that appeared real but were fake.[33]

---

[28] People v. Crabill, No. 23PDJ067, (Colo. O.P.D.J. Nov. 22, 2023).

[29] Hon. John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawy Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40, GA. ST. U. L. REV., 917, 927 (2024), https://readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=3275&context=gsulr.

[30] Pranshu Verma & Will Oremus, *These Lawyers Used ChatGPT to Save Time. They Got Fired and Fined.*, WASH. POST, https://www.washingtonpost.com/technology/2023/11/16/chatgpt-lawyer-fired-ai/ [https://perma.cc/TCU3-QLAW] (Nov. 16, 2023, 10:39 AM); see also, *Block v. Bramzon, No. B292129 (Cal. Ct. App. Jan. 22, 2021)*.

[31] John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawyers Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40 GA. ST. U. L. REV. 917, 928 (2023).

[32] United States v. Cohen, No. 18-cr-602 (S.D.N.Y. 2019).

[33] Andrew Zhang, *Michael Cohen's lawyer in hot water after citing court cases that don't exist*, POLITICO, (Dec. 12, 2023), https://www.politico.com/news/2023/12/12/michael-cohen-court-cases-00131435.

In *United States v. Pras Michel*,[34] the defendant, a former Fugees rapper, was convicted on multiple charges including conspiracy and funds to influence US politics. The respondent filed a motion for a fresh trial on the pretext that his former attorney had spoiled the defense by employing AI to draft closing arguments. Part of the defense argument for a new trial was based on the ineffective assistance of the prior counsel due to his financial stake in the AI company whose tools he deployed in closing arguments. Michel's new lawyer asserted that the AI tools generated frivolous arguments, damaging the defense because these arguments were deficient and prejudiced against the defense. The court concluded that the error did not prejudice the result of the case, hence the conviction was upheld.[35] This case raised significant ethical questions: Was the client informed of and to what extent did he agree to the employment of generative AI? What are the obligations to notify the judge of using generative AI? The case constitutes a warning to the attorneys that improper employment of generative AI may result in a breach of care, leading to a legal malpractice claim or lawsuit.

In another case,[36] the attorney submitted AI-generated response papers that contained fictional and flawed citations. While underscoring the significance of accuracy in legal documents, the court underlined the risks associated with AI-produced content without proper verification. Consequently, the court denied the motion for summary judgment, permitting the case to continue to factual disputes.[37]

In a recent case[38], plaintiff Iovino sued her former employer, Michael Stapleton Associates (MSA) for alleged whistleblower retaliation under federal law. The plaintiff claimed she was fired for reporting the defendant's contract with the US Department of State. The MSA counterargued that the petitioner had shared confidential information with the media and violated a non-disclosure agreement. The court addressed the plaintiff's objections to the protective order granted in favor of the MSA, which restricted certain discovery requests. The court overruled the plaintiff's objections, affirmed a protective order, and the plaintiff's attorneys were given a show-cause notice for presenting fictitious cases and made-up quotations.[39]

The Chief Justice of the US Supreme Court, John Roberts, in the annual judicial report of 2023, regarded hallucinations as a substantial impediment to AI integration in legal operations. Legal determinations often navigate gray areas where the application of human judgment is essential, so key actors in court cannot be fully replaced with machines.[40] Though the US courts are

[34] United States v. Michel, No. 19-cr-148 (D.D.C. 2023), https://www.courtlistener.com/docket/15511282/united-states-v-michel/.
[35] *Ibid*.
[36] In re Estate of Samuel, No. 2016-2501/A&B, 2024 N.Y. Slip Op. 24014 (Sur. Ct. Jan. 11, 2024), https://caselaw.findlaw.com/court/sur-s-crt-new-yor-kin-cou/115735333.html.
[37] *Ibid*.
[38] Iovino v. Michael Stapleton Associates, LTD., No. 5:2021cv00064 - Document 177 (W.D. Va. 2024).
[39] *Ibid*.
[40] Chief Justice John G. Roberts, Jr., *2023 Year-End Report on the Federal Judiciary,*6, (Dec. 31, 2023), https://www.uscourts.gov/news/2023/12/31/chief-justice-roberts-issues-2023-year-end-report.

sufficiently sensitized about legal hallucinations and their potential impacts, considering AI integration in legal operations. The courts urged lawyers to counter-verify AI-assisted filings. Still, they barely spoke about how and to what extent the fictitious authorities could harm the reputation of the judges and courts.

3.2. Case Law Development in Pakistan

In Pakistan, the integration of AI in legal operations is in its infancy. Interestingly, a judge in a recent case used ChatGPT-4 while adjudicating a civil lawsuit.[41] In his judgment, the judge explained how AI is transforming the future of legal adjudications. The court queried the LLM, ChatGPT, *"What are the principles for granting an injunction in a civil case in Pakistan?"* and compared the generated principles, which corresponded to the civil law (irreparable loss, balance of convenience, and prima facie case).[42] Nevertheless, the LLM produced three extra conditions for granting an injunction (good faith, public interest, and equitable consideration). These excessive conditions are not stipulated in the statutes for granting injunctions and may amount to hallucinations. The judge seems oblivious to the legal hallucinations and considers the additional conditions within the purview of statutory laws and the byproduct of the judicial precedents that evolved over the years.[43] In the instant case, the court overlooked statutory requirements where precision was paramount at the cost of securing the infallible character of the LLM. Therefore, the court declared the AI-generated results different in form but identical in substance, ignoring their inherent character of confabulation.[44]

In another case,[45] while granting pre-arrest bail to a juvenile, the same judge employed ChatGPT-4 to demonstrate how AI-powered solutions can help adjudication. As provided in the Order, the judge reported 18 responses[46] assigned to the GPT-4, which provided an interesting phenomenon for conceptualizing legal hallucinations: the conversation with the chatbot started with "Whether in Pakistan, a juvenile of 13 years is entitled to post-arrest bail?". In response to question No. 2, "Discuss it concerning section 83 of Pakistan penal code." the chatbot provided outdated information, children under 12 years are considered incapable of committing crimes. It failed to produce post-amended details, which the judge identified in question no. 3 that the age of sufficient understanding is now amended as 16 years.[47] In response to question no. 4, "In the above situation, if the offense is an attempt to commit rape, then what do you suggest? Option for bail request." the GPT summarized that "if the offense is an attempt to commit rape, the juvenile would not be entitled to bail as a matter of right under the Juvenile Justice System Act

[41] Muhammad Iqbal v. Zayad, (2023), CA 11 of 2023.
[42] *For details see,* Order 39, Section 94 (c) and (e) of the Code of Civil Procedure 1908 & Section 37(1) of the Specific Relief Act 1963.
[43] see (2014) PLD Sindh 268 (pak.); see also (2011) CLC 1866 (pak.).
[44] Bakht Munir, *Artificial Intelligence and Legal Decision-Making in the USA and Pakistan: A Critical Appreciation of Regulatory Frameworks* (Oct. 25, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4999590
[45] The State through Sameera Zulfiqar v. AM (a juvenile), FIR No.15/2023, dated 24.01.2023, Offence u/s 376(iii)/511PPC.
[46] For details, see ibid., pp. 6-15.
[47] *Ibid*., p. 7.

2018". The judge highlighted the next question "Section 6 of the Juvenile Justice System Act 2018 deals with bail".[48]

While responding to the task in question no. 6 to cite case laws where bail is granted in the same subject matter, the GPT sought pardon for not having access to case laws and databases.[49] Likewise, the chatbot refused to provide legal advice sought in question no. 11 and preferred to provide general information when the judge asked whether to grant bail in such an eventuality. The chatbot replied, "As an AI language model, I cannot provide legal advice, but I can provide some general information."[50] While responding to question no. 12, the GPT quoted the wrong provision, "497 of the Code of Criminal Procedure (Cr.PC)" for granting pre-arrest bail, supplemented with the advice to consult some qualified legal professional or refer to the updated laws and precedents. The judge corrected the relevant provision in question no. 13, "Please note that pre-arrest bail is considered under section 498 of Cr.PC."[51]

In response to question no. 15 regarding its inability to provide precedents on the matter under discussion, the GPT responded that it did not have real-time access to the Internet, was incapable of browsing case laws to interpret or analyze cases, and could not provide legal advice.[52] In question no. 16, the chatbot was assigned to cite some research articles on juvenile pre-arrest bail in rape or other cases in Pakistan's context. The GPT responded to the inability to directly quote or provide references to specific research articles due to not having access to external databases or internet browsing capability. The GPT suggested legal research databases such as Westlaw, LexisNexis, Pakistan Legal Research Database, JSTOR, and legal experts for assistance in providing relevant research articles and precedents.[53]

Considering all the discussions, the court observed that AI has great potential for the judicial system of Pakistan. The court realized the sensitivity of the matter and the disclaimer clause of the GPT emphasized further testing to exploit the potential of AI fully. Moreover, the judge sent a copy of the order to the Lahore High Court and the Law and Justice Commission of Pakistan for their perusal and consideration as a law reform proposal.[54] Interestingly, the judge regarded the chatbot's responses as impressive based on the correct appreciation of the laws, emphasizing the judiciary to rely on the AI solutions, avoiding any reference to the legal hallucinations that he encountered throughout with the chatbot.[55] In Pakistan, the experience of both cases exhibits that the integration of AI in legal adjudication is at its beginning. In both instances, the judge confronted excessive, incorrect, and outdated responses, though he remains oblivious to the legal hallucinations and their consequences which may end up in a miscarriage of justice. The Federal Judicial Academy of Pakistan provides judges across Pakistan with "*Judge-GPT" – an* AI-powered

[48] *Ibid*., p. 8.
[49] *Ibid*., p. 9.
[50] *Ibid*., p. 12.
[51] *Ibid*., p. 13.
[52] *Ibid*., p. 14.
[53] *Ibid*.
[54] *Ibid*., pp. 17-18.
[55] *Ibid*., p. 16.

solution – to assist the adjudication process. Neither the Pakistan Bar Council nor the Superior Judiciary of Pakistan have devised any conclusive ethics code to regulate the use of AI in legal operations. Moreover, the courts have yet to identify cases where lawyers using AI-powered solutions have submitted drafts supported by hallucinated references.

3.3. Case Law Development in the UK

The first reported case[56] in the UK where the court confronted AI hallucination was found when the cases cited by a litigant were not genuine but rather generated through AI solutions. In the instant case, Mrs. Harber failed to notify His Majesty's Revenue & Customs (HMRC) of her liability to capital gain tax following the disposal of her property. Consequently, she was issued a failure to notify penalty. She filed an appeal with the First-tier Tax Tribunal (FTT) against HMRC on the pretext of a reasonable excuse for her mental health and resulting ignorance of the law. She presented nine cases in which FTT sided with the taxpayer. However, the HMRC's representative asserted that the cases she presented were not identifiable. After the verification, the FTT concluded that the cases were not genuine, rather they were fabricated and generated through an AI tool like ChatGPT, though these cases were plausible but inaccurate. She confirmed that the cases were provided by a friend in a solicitor's office and could be AI-generated. The court disregarded the fabricated cases, and the appellant lost the appeal. The court opined that in addition to wasting time and other resources, attributing fake opinions to the judges and courts can damage their reputation, and harm the repute of any party wrongfully associated with illusionary conduct.[57]

3.4. Case Law Development in Australia

In a July 2024 hearing, a Melbourne lawyer was referred to investigation for presenting fabricated precedents in a family lawsuit. The attorney representing a husband provided the family court with a list of cases to support his plea. Neither the judge nor her associates could identify the enlisted cases. The lawyer confirmed that he used Leap, an AI-powered legal software specifically designed for legal use like Lexis+ AI, to prepare the list without verifying its accuracy and offered an unconditional apology. He paid the other party's solicitor for the costs of the thrown-away hearings. The court referred him to an investigation to appraise professional conduct issues based on the growing use of AI in legal operations. The family court has yet to issue guidelines on the use of AI. The Supreme Court of Victoria has already issued standards that the lawyers using AI should know the inherent limitations of these tools and how they work.[58]

Even though an AI model designed specifically for legal use can still create false or inaccurate information. AI solutions offer various means to validate their accuracy. For instance, 66,000 legal professionals are using Leap worldwide and it provides free verification through a human expert

---

[56] Harber v. Commissioners for His Majesty's Revenue and Customs, *[2023] UKFTT 1007 (TC)*, https://www.casemine.com/judgement/uk/65720f72cd29093de5347804.

[57] Burges Salmon, A cautionary tale of using AI in law; UK case finds that AI generated fake case law citations, UK, (Dec. 18, 2023), https://www.lexology.com/library/detail.aspx?g=18d97112-59a2-4513-af0f-bedc4bb594cc.

[58] Josh Taylor, *Melbourne lawyer referred to the complaints body after AI generated made-up case citations in family court,* (Oct. 10, 2024), https://www.theguardian.com/law/2024/oct/10/melbourne-lawyer-referred-to-complaints-body-after-ai-generated-made-up-case-citations-in-family-court-ntwnfb.

in the local laws, also known as human-in-the-loop. It is the lawyers' ethical obligation to verify the sources. Based on the request, the lawyer was provided with the correct information within four hours which he didn't utilize in court.[59] Before this, a group of Australian academics in November 2023, sought an apology for submitting AI-generated reports through *Google Bard,* now *Gemini,* against Big Four consultancy firms in submission to a parliamentary inquiry.[60]

3.5. Case Law Development in Canada

Likewise, In February 2024, a Canadian lawyer was referred to investigation for producing fictitious cases generated through ChatGPT in a child custody case in the Supreme Court of British Colombia. The attorney, Chong Ke, represented a father who wanted to take his children on a foreign trip. However, he was locked in a separation dispute with his wife. Chong Ke employed AI for precedents applicable to her client's circumstances. ChatGPT generated three responses and Key produced two of them in court. The opposing lawyer, however, could not trace those cases. Based on the confronted differences, Ke backtracked, maintaining the cases might be erroneous based on the AI-generated tool. She submitted an unconditional apology in the Court, having no intention to mislead the Court or the opposing counsel. The Court considered the submission of fake cases an abuse of process, which is equal to making false statements in the court and could lead to the miscarriage of justice. Her conduct is now under investigation by The Law Society of British Colombia, which issued guidelines on the appropriate use of AI in the delivery of legal services.[61]

3.6. Impacts of Hallucinations

Given the analysis of the cases, legal hallucinations may pose the following potential repercussions. In the first place, legal hallucinations impact lawyers by introducing inaccuracies to legal documents, damaging their integrity and credibility. It can breach ethical standards and professional responsibilities, leading to disciplinary actions and adding civil liabilities. In the second place, AI may augment legal services, but their hallucinations impact clients represented by the attorneys and may trigger distrust in the justice system. Inaccuracies in legal arguments can undermine their case, resulting in unfavorable judgments causing monetary losses or even wrongful convictions. In third place, legal hallucinations impact courts and judges, leading to miscarriage of justice. It diminishes the integrity of the judicial process, wasting time and resources to validate information and erode trust in the legal system.[62]

4. Response to AI Hallucinations

---

[59] Ibid.

[60] Henry Belot, *Australian academics apologize for false AI-generated allegations against big four consultancy firms* (Nov. 2, 2023), https://www.theguardian.com/business/2023/nov/02/australian-academics-apologise-for-false-ai-generated-allegations-against-big-four-consultancy-firms; *See also,* AI Hallucinations & Legal Pitfalls, (Sept. 17, 2024), https://www.madisonmarcus.com.au/news-media/areas-of-law/artificial-intelligence-law-areas-of-law/ai-hallucinations-legal-pitfalls/?cn-reloaded=1.

[61] Leyland Cecco, *Canada lawyer under fire for submitting fake cases created by AI chatbot,* (Feb. 29, 2024), https://www.theguardian.com/world/2024/feb/29/canada-lawyer-chatgpt-fake-cases-ai.

[62] John Doe, *Trust But Verify: Avoiding the Perils of Over-Reliance on AI in Legal Practice*, JD Supra (Dec. 1, 2024), https://www.jdsupra.com/legalnews/trust-but-verify-avoiding-the-perils-of-8176236/;

Considering the amplifying tendency towards AI in the legal province and its susceptibility to hallucination, a regulatory response in the form of patchwork has been evolving. The following segment examines responses to attorneys' irresponsible use of generative AI.

4.1. Judicial Responses

The increasing number of judges issuing AI orders varies in terms of breadth of coverage. Some judges prohibit the use of AI altogether, while some only prohibit it if lawyers do not verify accuracy; and some require submissions relating to the protection of confidential client information. We can categorize these responses into the following heads:

4.1.1. The Courts Requiring Confirmation on the Use of AI

After the New York federal court of show cause order in *Mata,* the first-ever reported case in which an attorney was caught using generative AI with fabricated outcomes, the Texas Court Judge Brantley Starr issued the first standing order governing the employment of generative AI.[63] Starr updated the individual practice rule by mandating a certificate about generative AI, which requires both the attorneys and the litigants to file a declaration in the court that no segment of the filing is drafted via generative AI, or if any segment is so drafted will be counter verified because these AI tools tend hallucinations and can provide biased information. In case of failure to file the required certificate, Starr's rule directed to strike such filing under Rule 11 of the Federal Rules of Civil Procedure irrespective of whether the draft or any portion thereof is AI-generated.[64]

Likewise, Judge Michael Baylson of the District Court of Pennsylvania issued a standing order regarding the disclosure of generative AI, requiring the attorneys to clarify where AI is used and to certify that each citation and reference has been verified.[65] Similarly, Magistrate Judge Gabriel Fuentes only mandated a certificate when a party actively uses generative AI, including disclosure about the filing and the specific tool employed.[66] Judge Scott Palk of Oklahoma issued the same

---

[63] Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 WL 4114965, at *1 (S.D.N.Y. June 22, 2023); Sara Merken, Wary Courts Confront AI Pitfalls as 2024 Promises More Disruption, REUTERS, https://www.reuters.com/legal/transactional/wary-courts-confront-ai-pitfalls-2024-promises-2023-12-27/; Shannon Capone Kirk, Emily A. Cobb & Amy Jane Longo, *Judges Guide Attorneys on AI Pitfalls with Standing Orders*, ROPES & GRAY (Aug. 2, 2023), Shannon Capone Kirk, Emily A. Cobb & Amy Jane Longo, Judges Guide Attorneys on AI Pitfalls with Standing Orders, ROPES & GRAY (Aug. 2, 2023), https://www.ropesgray.com/en/insights/alerts/2023/08/judges-guide-attorneys-on-ai-pitfalls-with-standing-orders.

[64] *Judge Brantley Starr – Judge Specific Requirements: Mandatory Certification Regarding Generative Artificial Intelligence*, U.S. DIST. CT. N. DIST. TEX., https://www.cit.uscourts.gov/sites/cit/files/Order%20on%20Artificial%20Intelligence.pdf.

[65] J. Michael M. Baylson, Standing Order Re: Artificial Intelligence ("AI") in Cases Assigned to Judge Baylson, (E.D. Pa. June 6, 2023), https://www.paed.uscourts.gov/sites/paed/files/documents/procedures/Standing%20Order%20Re%20Artificial%20Intelligence%206.6.pdf.

[66] Mag. J. Gabriel A. Fuentes, Standing Order for Civil Cases Before Magistrate Judge Fuentes, at 2 (N.D. Ill. May 31, 2023), https://www.ilnd.uscourts.gov/_assets/_documents/_forms/_judges/Fuentes/Standing%20Order%20For%20Civil%20Cases%20Before%20Judge%20Fuentes%20rev%27d%205-31-23%20(002).pdf.

standing order requiring disclosure about the use of AI and specific tools employed, coupled with a declaration about the accuracy of the draft and its supported citations.[67]

A New York Judge, Arun Subramanian, did not necessitate a disclosure but stressed that the attorneys must personally confirm the accuracy of the content before being presented to the court. The court ruled that the use of ChatGPT or other tools is prohibited unless the accuracy of these tools is personally confirmed.[68] On the other hand, New Jersey federal judge, Evelyn Padin, mandates the disclosure of the use of generative AI and certification that the accuracy of AI-generated content is confirmed under human supervision.[69] A District Judge of Hawaii, Leslie Kobayashi, directed that any party employing AI must disclose the use of AI along with the specific tool used and certify the authenticity of the generated contents, including citations. In case of default, the party will be held accountable under Rule 11, which may lead to the imposition of sanctions.[70] The US Magistrate Judge Jeffrey Cole of Illinois while adopting the standing order for the use of generative AI, requiring both disclosure and certification. The court declared that generative AI, by producing fabricated and inaccurate citations, compromises the court's mission to ascertain truth.[71]

In addition to the trial courts, other US federal judges have followed Judge Starr's pattern for governing the use of AI. For instance, the Bankruptcy Court of Texas requires that if someone uses generative AI while preparing a filing, they must ensure the accuracy of the generated text through reliable means, including conventional legal databases and print reports.[72] In the appellate courts, the US Court of Appeals for the Fifth Circuit was the first to give notice of the proposed rule governing the employment of generative AI. The court proposed an amendment to Fifth Circuit Rule 32.3 to add language addressing AI use to its existing certificate of compliance, which includes a certificate of whether generative AI was used, its extent, and its accuracy approval by a human.[73] Likewise, Juge Roy Ferguson of the 394th District Court in Texas was the first state court to pass a standing order governing the employment of generative AI. The order

[67] J. Scott L. Palk, Chambers of United States District Judge, Disclosure and Certification Requirements – Generative Artificial Intelligence, https://perma.cc/VYZ8-XNGH.

[68] J. Arun Subramanian, United States District Court Southern District of New York, Individual Practices in Civil Cases, at 7 (2023), https://perma.cc/SNN5-N6HR.

[69] Judge Evelyn Padin's General Pretrial and Trial Procedures 2 (2023), https://perma.cc/M6RY-FVGP.

[70] J. Leslie E. Kobayashi, Chambers of United States District Judge, Disclosure and Certification Requirements – Generative Artificial Intelligence, https://perma.cc/Z63A-VSQX.

[71] Mag. J. Jeffrey Cole, United States District Court for the Northern District of Illinois, The Use of "Artificial Intelligence" in the Preparation of Documents Filed Before This Court, https://www.ilnd.uscourts.gov/_assets/_documents/_forms/_judges/Cole/Artificial%20Intelligence%20standing%20order.pdf.

[72] United States Bankruptcy Court for the Northern District of Texas, General Order 2023-03, Pleadings Using Generative Artificial Intelligence (June 21, 2023), https://perma.cc/JQ6Y-THKV.

[73] Jacqueline Thomsen, Lawyers Must Certify AI Review Under Fifth Circuit Proposal, BLOOMBERG L. (Nov. 21, 2023, 6:26 PM), https://news.bloomberglaw.com/us-law-week/lawyers-must-certify-ai-review-under-fifth-circuit-proposal; see also, https://www.ca5.uscourts.gov/docs/default-source/default-document-library/public-comment-local-rule-32-3-and-form-6.

mandated the filers to certify that all the generative content is substantiated as accurate via conventional legal methods by an attorney licensed to practice law in Texas.[74]

4.1.2. Non-Disclosure of Confidential Information

A federal Judge, Stephen Alexander Vaden of the United States Court of International Trade, issued an 'Order on AI' showing concerns about privacy invasion that these tools learn from users' interaction and cannot differentiate between confidential and non-confidential information. Hence, Judge Vaden mandated two things with AI-generated filings: A disclosure notice regarding the tool employed along with the segment generated and a declaration that the use of AI has not disclosed any confidential information to an unauthorized person.[75] Likewise, the Bankruptcy Court of Oklahoma, while quoting Judge Starr, mandated disclosure about the AI tool, the details of the specific portion for which generative AI was employed, a certificate of accuracy checking, and to confirm that generative AI has not caused the disclosure of any confidential information to any unauthorized party.[76]

4.1.3. The Courts Prohibiting the Use of AI Solutions

The US District Judge of Montana, Donald Molloy, prohibited the employment of generative AI software like ChatGPT.[77] Judge Michael Newman of Ohio prohibited the use of generative AI and warned of the sanctions that might be imposed for using AI, including monetary, contempt, and dismissal of the suit. However, the court allows information collection from legal search engines like LexisNexis and Westlaw and common search engines like Google.[78] Similarly, Judge Stephen Clark of Missouri banned the use of generative AI.[79]

The courts' responses to the use of AI vary across the US necessitating the attorneys to double-check each court's policy on the use of AI before filing any submission to avoid any potential complications. As discussed, some courts allow the employment of AI subject to the disclosure of its use, the tool so employed, the extent of its assistance, and the confirmation of its accuracy. In addition to these standards, some courts require the confirmation that the employment of AI has not disclosed clients' confidential information to any unauthorized person. In contrast, some courts prohibited the use of AI altogether.

4.2. Other Responses

---

[74] District Court for the 394th Judicial District of Texas, Standing Order Regarding Use of Artificial Intelligence (June 9, 2023), https://edrm.net/wp-content/uploads/2024/04/Judge-Roy-Ferguson.pdf.

[75] Hon. Stephen Alexander Vaden, Order on Artificial Intelligence, 1 (Ct. Int'l Trade June 8, 2023), https://www.cit.uscourts.gov/sites/cit/files/Order%20on%20Artificial%20Intelligence.pdf.

[76] United States Bankruptcy Court for the Western District of Oklahoma, General Order 23-01, Pleadings Using Generative Artificial Intelligence (July 25, 2023), https://www.okwb.uscourts.gov/sites/okwb/files/GenOrder23-01.pdf.

[77] *Belenzon v. Paws Up Ranch, LLC, No. CV 23-69-M-DWM, 2023 U.S. Dist. LEXIS 123020, at 1 (D. Mont. June 22, 2023), https://casetext.com/case/belenzon-v-paws-up-ranch-llc*.

[78] Hon. Michael J. Newnan, United States District Court for the Southern District of Ohio, Standing Order Governing Civil Cases, at 11 (Dec. 18, 2023), https://perma.cc/V6P6-BSRZ.

[79] Self-Represented Litigants (SRL), U.S. Dist. Ct. E. Dist. Mo.: Hon. Stephen R. Clark, C.J. & Nathan M. Graves, Clerk of Ct., https://www.moed.uscourts.gov/self-represented-litigants-srl.

Following the Mata case ruling, the legal community is now more aware of using generative AI tools, necessitating policy guidelines for governing AI in legal operations. The policy should sensitize legal professionals about the appropriate use of AI and its ethical concerns. AI should be employed only to assist and augment legal services, but not at the cost of lawyers' subjective judgment and expertise. Moreover, attorneys should be held responsible for validating the accuracy of the generated contents. While employing AI in legal services, clients should also be taken into confidence. Attorneys should remain current about the emerging trends in AI as their ethical duty of technological competence.

MIT convened the first task forces in response to Mata's case to ensure responsible use of generative AI.[80] The State Bar of Texas initiated a task force to explore the proper employment of AI in legal services. The task force aimed to ensure that technological advancement served the community without compromising values central to the legal community. The Texas Task Force made numerous recommendations to the state bar, including technological education and ethical use of AI.[81] The New York Bar Association also declared its own AI task force to appraise the impacts of evolving technology on the legal profession and society.[82] The American Bar Association (ABA) announced the formation of a national task force to assess the risks of AI on the legal profession, including data privacy, disinformation, and cybersecurity. Further to examine AI governance, AI in legal education, and AI in access to justice. To address the impacts and ethical concerns of AI and provide insights on the trustworthy and responsible use of AI.[83]

In addition to the task forces, two ethics bodies have responded to the issue of AI. The Board of Governors Review Committee of the Florida Bar considered an advisory opinion on the use of AI after an inquiry on AI tools. The committee issued a proposed advisory opinion to address issues that the attorneys employing AI must take reasonable steps to safeguard clients' privacy information, a reasonable oversight on the use of generative AI, and lawyers must not entrust their subjective judgment to generative AI. The proposed opinion also demands lawyers to charge only a reasonable fee and should not overly charge their clients for using AI. Lawyers may publicize the employment of generative AI but cannot claim its authority over others unless the same is objectively verifiable. Since generative AI is still in its beginning, the existing ethical concerns should not be treated as final.[84]

---

[80] Dazza Greenwood, Task Force on Responsible Use of Generative AI for Law, MIT Computational Law Report (Feb. 28, 2023), https://law.mit.edu/pub/generative-ai-responsible-use-for-law/release/9.

[81] State Bar of Tex., Taskforce for Responsible AI in the Law (Trail) 2–3 (2023), https://www.texasbar.com/AM/Template.cfm?Template=/CM/ContentDisplay.cfm&ContentID=61655.

[82] Richard Lewis, What the NYSBA AI Task Force Hopes to Achieve for Law Practice, BLOOMBERG L. (July 31, 2023), https://news.bloomberglaw.com/us-law-week/what-the-nysba-ai-task-force-hopes-to-achieve-for-law-practice.

[83] ABA Forms Task Force to Study Impact of Artificial Intelligence on the Legal Profession, AM. BAR ASS'N (Aug. 28, 2023), https://www.americanbar.org/news/abanews/aba-news-archives/2023/08/aba-task-force-impact-of-ai/.

[84] Proposed Advisory Opinion 24-1 Regarding Lawyers' Use of Generative Artificial Intelligence – Official Notice, FLA. BAR (Nov. 13, 2023), https://www.floridabar.org/the-florida-bar-news/proposed-advisory-opinion-24-1-regarding-lawyers-use-of-generative-artificial-intelligence-official-notice/.

The State Bar of California necessitated the governance of generative AI. Its Committee on Professional Responsibility and Conduct (COPRAC) provided recommendations on and stipulated practical guidance on the use of generative AI. It examines how generative AI impacts professional responsibility obligations, including confidentiality, competence, supervision, and charging only a reasonable fee.[85]

Additionally, the state of Michigan is accredited to be the first to issue a Judicial Ethics Opinion, addressing judges' ethical obligation concerning generative AI that judicial officers must keep up with technological advancements including AI. It further says that with the proliferation of AI, the judges must comprehend the legal, regulatory, and ethical challenges of AI and consistently appraise how they or parties before them are employing AI in their docket.[86]

In Pakistan, the National Artificial Intelligence Policy, 2022 is launched with the proposed establishment of an AI regulatory directorate to ensure the ethical and responsible use of AI.[87] However, there is no reference to dealing with the emerging issues of AI hallucinations in decision-making. Notably, the Federal Judicial Academy of Pakistan facilitated judges across Pakistan with the *Judge-GPT* AI tool to assist the decision-making process, without providing guidelines about its probabilistic nature that could lead to plausible but inaccurate responses. In critical areas like health, finance, and law, where accuracy is paramount, fabricated outcomes can cause irreparable loss. In legal services, fictitious precedents could cause a miscarriage of justice. Neither the government, bar, bench, nor law firms have established definite standards on the rapidly evolving issue of hallucinations. Necessitating the establishment of a task force, comprising experts from the academia, government, judiciary, and tech developers, to devise an exclusive policy to deal with AI in legal practices and its ethical challenges.

In the UK, Artificial Intelligence (AI) Guidance for Judicial Office Holders, 2023, offers comprehensive guidelines about the responsible employment of AI. It underscores the limitations and capabilities of AI and urges its conformity with the judiciary's overreaching obligation to protect the integrity of the administration of justice. It applies to all the courts and tribunals across the UK and provides the following guidelines for the responsible use of AI: the AI chatbot produces results based on the prompts they receive, the data they are trained, the information available on the internet, and may generate a plausible but inaccurate response. Confidentiality and privacy are another concern. The public chatbot retains every prompt and information, which may be utilized in responding to other users, invading data privacy. Likewise, the accuracy of the responses must be confirmed before being used or relied upon. The AI tools may cause fabricated citations, cases, and quotes, or may refer to legal text that doesn't exist. Hence, these tools cannot be left unaccountable. It further provides for biases, security, responsibility, and potential employment of AI by other users. The draft exemplifies the positive use of AI such as its capabilities to summarize large legal text, prepare presentations, and perform administrative

[85] Memorandum from the Comm. on Pro. Resp. & Conduct to Members, Bd. of Tr. Sitting as the Regul. and Discipline Comm. 1 (Nov. 16, 2023), https://aboutblaw.com/bbpZ.
[86] State Bar of Mich., Ethics Op. JI-155 (2023), https://perma.cc/C58T-GCLX.
[87] Ministry of Information Technology and Telecommunication, National AI Policy Consultation Draft V1 (2022), https://moitt.gov.pk/SiteImage/Misc/files/National%20AI%20Policy%20Consultation%20Draft%20V1.pdf.

tasks like drafting emails and memos. However, the draft discourages conducting legal research on AI tools that cannot be independently counter-verified and legal analysis because the current tools are incapable of producing convincing reasoning or analysis.[88]

The Law Society of England and Wales also provided an AI strategy focusing on the following three long-term outcomes: innovation, to benefit both firms and clients; impact, to have an effective regulatory landscape; and integrity, to ensure the responsible and ethical employment of AI to advance the rule of law and access to justice. It embraces endeavors to ensure that the legal system operates impartially, safeguards individual rights, and advances the cause of justice, including the protection of the rights of marginalized communities, addressing prejudices, and striving to ensure that the legal system upholds principles of justice for every member of society.[89]

In Australia, the legal profession regulators from across the three uniform law states have jointly issued a statement to guide legal professionals in their ethical and responsible use of AI: the Law Society of NSW, the Legal Practice Board of Western Australia, and the Victorian Legal Service Board and Commissioner have established common principles to protect the client from risk, technology is employed for their benefits, and uphold the principles of justice. The following are the key points of the statement: while enjoying the assistance of AI, lawyers are obliged to provide accurate legal information, and it is not the duty of the AI tool being employed. The practitioners must understand AI, its capabilities, and the limitations of LLMs. This statement helps the lawyers understand the regulators' expectations when they employ AI to assist them in providing legal services. The regulators will frequently review and update their guidance on AI as it continues to evolve. While using AI, legal professionals must maintain high ethical standards and rules of conduct, including upholding clients' confidentiality, advising their clients, competent and diligent provision of legal services, charging a reasonable, fair, and proportionate fee, transparency, and limiting the use of AI.[90]

The Canadian Judicial Council (CJC) issued Guidelines for the Use of AI in Canadian Courts, 2024, which provides a framework for the responsible use of AI in judicial processes. It underlines upholding judicial independence, adhering to the core values and ethical standards, and ensuring information security, transparency, and accountability in AI-generated content. It underscores the significance of regular impact assessments, sensitizing, and user support for judges. It aims to strike a balance between innovation and caution, ensuring that AI advances the efficiency of legal services without compromising the integrity of the judicial system. The guidelines are broadly categorized into the following seven heads. First, protection of judicial independence, restricting the parliament's authority to empower a state agency from the legislative and judicial branches to oversee the use of AI by and before courts. Where the government moves forward with legislation to regulate AI, the independence of the judiciary must be preserved. Second, judges'

[88] Judicial Office, Artificial Intelligence (AI) Guidance for Judicial Office Holders (2023), https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf.

[89] The Law Society, Artificial Intelligence (AI) Strategy (2023), https://www.lawsociety.org.uk/topics/ai-and-lawtech/artificial-intelligence-ai-strategy.

[90] Legal Services Board of Victoria, Statement on the Use of Artificial Intelligence in Australian Legal Practice (2023), https://www.lsbc.vic.gov.au/news-updates/news/statement-use-artificial-intelligence-australian-legal-practice.

employment of AI must adhere to the core values and ethical rules, including integrity, competence, impartiality, transparency, fairness, and accessibility to justice. Third, regards aspects like privacy and intellectual property, creating an equilibrium between safety and accuracy. Fourth, strictly adhere to the information security standards through robust information and cyber security programs. Fifth, the AI tools must provide reasons and explanations for their decision-making and generative outcomes. Sixth, to keep regular track of the use of AI considering judicial independence, security, privacy, access to justice, and the court's reputation. Seventh, user support and education, including judges training which is a prerequisite for upholding and maintaining judicial independence and technical support for AI integration in the administration of justice. The seven points outlined by the CJC reaffirm that AI should not be employed without a comprehensive understanding of the best practices for integrating technology.[91]

To conclude the above responses, the judges, bar, and law firms contribute to developing AI rules, but their contribution is a patchwork. The courts' responses create confusion even in the patchwork: certain courts proscribed the employment of generative AI, few require disclosure and certification, while some do not. Some judges are concerned about data confidentiality. Hence, attorneys should stay vigilant of technological advancement considering the applicable and often diverging court rules. The growing tendency of AI in legal operations necessitates an exclusive national policy governing the use of generative AI and its ethical challenges in the legal province.

The most striking question is how to overcome hallucinations in legal operations. Legal hallucinations are the byproduct of many contingencies and could be addressed accordingly. The following segment explains how to curtail if could not completely remove, the issue of hallucinations from AI-powered solutions.

5. Debugging AI Hallucinations in Legal Operations

One of the main concerns in AI legal practice is dealing with AI hallucinations. Considering its probabilistic nature and its susceptibility towards fabricated responses, the AI hallucination mitigating techniques can be broadly divided into the following two heads:

5.1. Recommendations for the Developers

The use of high-quality training data helps diminish the prospects of hallucinations. The first stage that leads to the likelihood of hallucinations is the lack of accuracy and reliability of datasets. Hence, hallucinations are inversely proportional to the accuracy and consistency of datasets. Hallucinations tend to decline as the precision and trustworthiness of the training data enhance so using data templates or structured data formats is advisable. Improving the quality of the training sample and subsequent testing of the generative data can help reduce the possibility of hallucinations. Clearly outlining what AI is tasked to do can generate focused and appropriate outcomes. Putting restrictions on the AI's responses can help improve the performance and reliability of the LLMs.

[91] Canadian Judicial Council, Guidelines for the Use of Artificial Intelligence in Canadian Courts (2024), https://cjc-ccm.ca/sites/default/files/documents/2024/AI%20Guidelines%20-%20FINAL%20-%202024-09%20-%20EN.pdf.

By applying modern artificial neural network architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs), the rate of hallucinations can be significantly controlled: CNNs are useful for comprehending the context and structure of legal documents because they are exceptional to identify spatial hierarchies in data. By deploying CNNs to the legal data, LLMs can help realize the complexities and intricacies of legal language, diminishing the prospects of fictional outcomes. Likewise, LSTMs can help improve sequence prediction because they are designed to retain long sequential data, making them best for dealing with extensive legal documents, preserving context for an extended period, and allowing the LLMs to generate precise and relevant responses. LSTMs address vanishing and gradient problems encountered in other networks.[92] A hybrid of both these architectures can help design more robust models: CNNs for extracting features from the legal data while LSTMs for handling the sequential nature of legal documents, leading to a comprehensive understanding and producing legal text, reducing the likelihood of hallucinations. Extensive training of these models on specific legal data helps advance the accuracy of these models when they are exposed to voluminous legal text to learn different terminologies and contexts in the legal province, cutting errors and hallucinations.

Continuous model improvement, consistent updates, and advancements in AI models can help reduce the prospects of hallucinations, so regular appraisal and improvements of these models are inevitable. Human oversight is a significant tool to control the prospects of hallucinations. LLMs can be trained enough to overcome potential hallucinations by monitoring and correcting AI's responses.[93] AI-generated content should also undergo regression analysis before being presented or relied upon. Further, an ethical supervisor is advisable for the algorithms to monitor and impose ethical restrictions on the use of AI based on the idea that the former must have higher standards than the latter.[94] Humans in the loop are also recommended to review and correct AI responses with a focus to train AI intelligently not to repeat a fictitious outcome, and the end user may get the generated content counter-verified before relying on it.

Fine-tuning models for specific legal tasks may also reduce the likelihood of producing inaccurate responses by making small adjustments to the model's parameters, based on the existing knowledge of a model to learn new tasks. Implementation of robust evolution of metrics to frequently assess and address hallucination rates is also recommended. Another way to moderate the rate of hallucinations is to design a self-explanatory AI model, which can provide explanations and reasons for its decision-making process. This proposed model can help legal professionals

---

[92] Sepp Hochreiter et al., *Long Short-Term Memory*, 9, Neural Computation, (8): 1735–1780 (1997), https://dl.acm.org/doi/10.1162/neco.1997.9.8.1735; *see also, What is LSTM? Introduction to Long Short-Term Memory, Analytics Vidhya*, (Oct. 1, 2024), https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/.

[93] What are AI hallucinations? IBM, https://www.ibm.com/topics/ai-hallucinations.

[94] Vadim Perov and Nina Perova, *AI Hallucinations: Is "Artificial Evil" Possible?,* USBEREIT, (2024), https://ieeexplore.ieee.org/document/10584048; *See also,* Amitai Etzioni and Oren Etzioni, *AI assisted ethics*, 18, Ethics Inf. Tech., (2016), https://link.springer.com/article/10.1007/s10676-016-9400-6.

identify potential hallucinations and appraise the reliability of AI content, resulting in more transparent and accountable AI systems.

Though these suggestions can significantly contribute to curtailing the rate of hallucinations, they cannot be completely overcome. LLMs operating on probabilistic algorithms attempt to predict or foresee the next word in a sequence by considering the prospect of various possible words that may lead to a conceivable but inaccurate response.[95] These processes by their very nature can lead to inaccuracies and potential hallucinations. Despite the widespread and high-quality training data, no dataset can cover every eventuality due to the complexity and context-dependent nature of languages.[96] Unlike conventional search engines, these models are designed to be creative and capable of generating diverse and stimulating results. Nevertheless, precision can compromise creativity, creating a challenge in maintaining an equilibrium between the two. The inherent limitations of the current machine learning algorithm create prospects of hallucinations in LLMs when generalizing from training data to unanticipated data.[97]

5.2. Recommendations for Legal Professionals

Given the above discussion, AI models still have the potential to produce fabricated responses owing to their probabilistic nature. Precision is highly valued in legal operations, so it is highly recommended that legal professionals counter-verify AI-generated content and citations against reliable sources. AI solutions should be used as a supplement to augment legal services rather than a substitute. Legal professionals must stay abreast of the limitations, common kinds of hallucinations and errors specific to legal context, and pitfalls of AI tools.

Prompt skilling can substantially mitigate the prospects of hallucinations. By crafting precise and effective commands, the AI models recognize exactly what is being queried. Accuracy and comprehension of these models can be further improved by employing the following techniques: prompt chaining, which breaks down a long and complex proposition into simple and sequential inputs. Employing multimodal or a diversity of prompts can help enhance AI comprehension. Consistent appraisal and feedback significantly contribute to refining AI models. Another advisable solution is to pass the AI-generated content through robust quality control by establishing review protocols for AI responses, including multiple layers of review: cross-referencing AI information with conventional databases, consultation with colleagues, peer reviews, human oversight, other rounds of fact-checking, and maintaining a healthy skepticism towards AI content.[98]

---

[95] Major research into 'hallucinating' generative models advances reliability of artificial intelligence, University of Oxford, (Jun 20, 2024), https://www.ox.ac.uk/news/2024-06-20-major-research-hallucinating-generative-models-advances-reliability-artificial.

[96] Changlong Wu at al., *No Free Lunch: Fundamental Limits of Learning Non-Hallucinating Generative Models*, CSol, Preprint under review, (2024), https://www.cs.purdue.edu/homes/spa/papers/hallucination_preprint.pdf.

[97] Minhyeok Lee, *A Mathematical Investigation of Hallucination and Creativity in GPT Models*, 11 Mathematics 2320 (2023), https://doi.org/10.3390/math11102320.

[98] AI Hallucinations: Legal Information Risks, Attorneys Media, https://attorneys.media/ai-hallucinations-legal-information-risks/.

Legal professionals should regularly update their knowledge of AI innovations and best practices by participating in seminars and training sessions based on the efficient and ethical employment of AI in legal services. A regular audit of the AI tools is needed to ensure compliance with the approved standards and relevant laws, necessitating calls for a regulatory framework.[99] A loop among legal researchers, practitioners, and AI developers is highly recommended for designing more trustworthy AI models. The developers should provide guidelines and training to their users on the effective employment of AI in legal services. Legal practitioners should stay connected with the AI service providers to report flaws and propose improvements, which can significantly contribute to refining these models by not repeating a specific hallucination.

6. Conclusion

AI-driven models are revolutionizing legal operations, simultaneously creating inherent challenges in navigating legal landscapes. Despite high-quality training and data optimization, LLMs are susceptible to hallucinations. This intrinsic problem is the outcome of their functional modalities: considering their probabilistic nature, the LLMs calculate the possibility of a particular word following in a sequence. While training the data, these models learn patterns, structures, and correlations between the words. These models follow a sequence based on the assigned probability of each word. These models depend on the context provided by the preceding data and complete the sequence of words by generating the most probable content comprehended in their co-relationship. In the given scenario, these models cannot verify the truthfulness and relevance of the context, hence the required outcomes might be plausible but imprecise or fictitious. The models are only concerned with a high probability of words next in sequence. These models, however, cannot authenticate the trustworthiness of their generated content, thus adding a disclaimer that AI-generated content may be inaccurate is feasible to exonerate civil law liability. It shifts the onus to the end user to counter-verify the generated content otherwise face the music.

Imposing limitations on the training data can help narrow the likelihood of false or fabricated content at the cost of creativity. Conversely, the models trained on widespread data without such confines may produce more novel outcomes. In sensitive fields like finance, healthcare, and law where precision is paramount, utmost care to avoid hallucinations is required, though at the expense of novelty. An equilibrium between hallucinations and creativity can help design a more robust and versatile model, capable of addressing complex tasks with enhanced performance, leaving the end user with ultimate liability to corroborate the generated content before relying on it.

Legal professionals are swiftly integrating AI systems into their legal provinces without fully realizing their probabilistic nature which could lead to fabricated outcomes, affecting the cause of justice. Given their utility despite their unpredictable nature, these AI systems can be referred to as necessary evil. They are unavoidable owing to the unparalleled services they offer, but their

[99] Kiara Brunel Fink, *AI Hallucinations in Legal Practice: Risks, Real Cases, and Solutions,* Mondaq, https://www.mondaq.com/new-technology/1540712/ai-hallucinations-in-legal-practice-risks-real-cases-and-solutions.

irresponsible employment can transpire into malpractice, compromising the reputation of lawyers, and creating financial liability. Even if judges and lawyers are reluctant to use AI, they still need to learn AI. Since AI can go wrong, legal professionals are under obligation to act as guardians of the legal system to correct their abuses.[100]

Transparency and accountability could help moderate the probability of hallucinations. The AI enterprises must be transparent about conceivable errors, including accountability measures and setting up expectations for clients where the AI-produced content leads to issues. Bar Associations such as California, Florida, and New York have published guidelines for the trustworthy use of AI in legal operations.[101] More than 25 US Federal Judges passed standing orders requiring lawyers to reveal and circumscribe the use of IA in their courtrooms.[102] The judges' directions to the attorneys to certify the use and accuracy of AI in their briefs are motivated by the ethical challenges posed by AI and the significance of the precision of documents submitted in the court.

By incorporating a disclaimer about the accuracy of the generated content, the AI developers exculpate their liability for disseminating fictitious content. However, AI responses based on erroneous opinions could damage the reputation of judges, courts, or parties implicated in fictional conduct. The greatest risk lies on the part of the legal user who may not and arguably should not be able to escape liability for over-reliance on AI. Although these tools are still in their developmental stages and evolving towards maturity, the judiciary and legal community must determine the acceptable extent of their fabricated responses, necessitating the establishment of policy guidelines. Regardless of their legal liability, AI enterprises are responsible for providing trustworthy and reliable services. They must adhere to ethical and legal standards, confirming their models do not create harmful or misleading responses.

To have confidence in the AI solutions, a shared liability clause in user agreements should be incorporated, clearly outlining the responsibility of both the users and the service providers and demonstrating the extent of their liability in cases where AI hallucinations cause harm. For instance, the European Union (EU) has recently initiated a Product Liability Directive (PLD), placing obligations on AI tool developers, suppliers, and other entities for providing defective products, including AI software. So, the manufacturers can be held accountable for the harm

---

[100] David Rubenstein, *2024 Selected Topics and Miscellany CLE,* Washburn University School of Law, Presentation (June 13, 2024), https://www.washburnlaw.edu/about/community/cle/_files/selected-topics-schedule.pdf.

[101] David Alexander, New York State Bar Association Task Force To Address Emerging Policy Challenges Related to Artificial Intelligence, N.Y. St. Bar Ass'n (July 17, 2023), https://nysba.org/new-york-state-bar-association-task-force-to-address-emerging-policy-challenges-related-to-artificial-intelligence/; *See also*, Report and Recommendations of the New York State Bar Association Task Force on Artificial Intelligence, N.Y. St. Bar Ass'n (April 2024), https://nysba.org/app/uploads/2022/03/2024-April-Report-and-Recommendations-of-the-Task-Force-on-Artificial-Intelligence.pdf; The State Bar of California, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law, State Bar of Cal. (2023), https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf; The Florida Bar, Florida Bar Ethics Opinion, Technical Report 24-1, Fla. Bar (2024), https://www.floridabar.org/etopinions/opinion-24-1/.

[102] Law360, Tracking Federal Judge Orders on Artificial Intelligence (2024), Pulse (law360), https://www.law360.com/pulse/ai-tracker.

caused by defects in their AI solutions, and the injured party is not even required to prove fault or negligence.[103] In addition to PLD, the EU is considering the AI Liability Directive (AILD), to address risks posed by AI tools by introducing a fault-based civil liability regime, which would require proving the developer's fault or negligence where AI solutions cause harm. These directives are part of the EU's comprehensive efforts to regulate AI, offering users legal pathways to seek compensation for any damage caused by AI tools.[104]

[103] Kennedys Law, *A New Liability Framework for Products and AI*, https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/.

[104] Giskard, *AI Liability in the EU: Business Guide to Product (PLD) and AI Liability Directives (AILD),* https://www.giskard.ai/knowledge/ai-liability-in-the-eu-business-guide-to-product-pld-and-ai-liability-directives-aild; Kennedys Law, A New Liability Framework for Products and AI, https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/.