

Hallucinated Authority: AI Citations as Reckless Misrepresentation

CHENG-CHI (KIRIN) CHANG*

When courts sanctioned the attorneys in Mata v. Avianca, Inc. for submitting artificial intelligence (AI)-generated citations to nonexistent cases, they reached for the vocabulary of negligence. That characterization is wrong. Large language models do not forget cases the way attorneys do; they generate citations that look real but correspond to nothing in any reporter, and they do so, as studies have since documented, at rates between fifty-eight percent to eighty-eight percent on legal queries. The tools' general propensity to fabricate was publicly known by 2023, before most of the incidents now filling the reporters. An attorney who submits AI output without verification does not merely fail a diligence standard. They submit a representation to a court while consciously declining to establish any basis for its accuracy, satisfying the elements of reckless misrepresentation under Restatement (Second) of Torts § 526(b)–(c). The cases that have followed Mata reach instinctively for language that exceeds negligence, calling citations “bogus” and monetary sanctions “insufficient deterrents,” and some have begun to reach for the word “reckless” without grounding that intuition in a cross-regime doctrinal framework. This Essay supplies that framework.

This Essay makes three original contributions. First, it introduces the concept of “structurally predictable error” to explain why AI hallucination is categorically different from human memory failure: the error follows from the tool’s architecture rather than from individual cognitive lapse, its rate is publicly documented, and a reasonable practitioner in the field either knows or should know its hallucination rate before using the tool. Second, it builds the doctrinal case for reckless misrepresentation by connecting Section 526(b)–(c) to the philosophy of AI output, arguing that an attorney who relies on a tool architecturally indifferent to the distinction between truth and falsity adopts that indifference as their own representational posture toward the court. Third, it proposes a graduated three-tier accountability framework that calibrates sanctions to actual culpability, preserving space for good-faith verification efforts and the access-to-justice benefits that AI tools offer to under-resourced practitioners, while refusing to treat willful blindness toward a known category of fabrication as mere carelessness.

* Cheng-chi (Kirin) Chang (張正麒) is a Research Fellow at the U.S.-Asia Law Institute of NYU School of Law, a fellow at NYU Law’s Information Law Institute, and an Affiliate Research Fellow of the AI and the Future of Work Program at Emory University School of Law, where he previously served as Associate Director and Academic Fellow. I am grateful to Katherine A. Wilhelm, Ifeoma Ajunwa, and the many friends and colleagues whose generosity and encouragement have meant so much to me, in this work and well beyond it. I also thank Sylvia (Xi) Chen, Lexi Detrick, Emma Marszalek, and the other Editors of *The Georgetown Law Journal* for their care and dedication in bringing this Essay to publication. Any errors or oversights are my sole responsibility.

TABLE OF CONTENTS

INTRODUCTION	41
I. THE ANATOMY OF AI HALLUCINATION	43
A. GENERATING LEGAL TEXT.....	43
B. MEASURING HALLUCINATION RATES	44
C. THE CONCEPT OF STRUCTURAL PREDICTABILITY	45
II. INADEQUACIES OF THE CURRENT LEGAL FRAMEWORK	46
A. RULE 11 AND OBJECTIVE REASONABLENESS	46
B. MODEL RULE 3.3 AND THE DUTY OF CANDOR.....	47
C. STATE BAR GUIDANCE AND PROFESSIONAL COMPETENCE	47
III. THE CASE FOR RECKLESS MISREPRESENTATION.....	48
A. THE ARCHITECTURE OF RECKLESSNESS	48
B. APPLYING THE RECKLESSNESS STANDARD	49
C. EPISTEMOLOGICAL INDIFFERENCE	50
D. THE STRUCTURAL PREDICTABILITY TEST	50
IV. A GRADUATED ACCOUNTABILITY FRAMEWORK.....	51
A. ACCESS-TO-JUSTICE CONCERNS.....	51
B. A THREE-TIER ACCOUNTABILITY MODEL	52
C. MANDATORY DISCLOSURE OBLIGATIONS.....	54
D. TECHNOLOGICAL SAFEGUARDS AND RAG MODELS.....	56
E. COMPARATIVE INTERNATIONAL DEVELOPMENTS.....	57
CONCLUSION.....	58

INTRODUCTION

In March 2023, attorney Peter LoDuca filed an “Affirmation in Opposition” in the Southern District of New York that cited, among other cases, *Varghese v. China Southern Airlines Co., Ltd.*, *Shaboon v. Egyptair*, and *Martinez v. Delta Airlines, Inc.*¹ None of those cases existed.² A generative artificial intelligence (AI) tool called ChatGPT had fabricated them, and LoDuca submitted them without verification.³ When the court ordered him to produce the opinions, he submitted fabricated ones instead.⁴ Judge P. Kevin Castel called the submissions “bogus judicial decisions, with bogus quotes and bogus internal citations” and sanctioned LoDuca, the filing’s author Steven Schwartz, and their firm a total of \$5,000.⁵

That incident was not an outlier. In January 2024, the Second Circuit referred attorney Jae S. Lee to its Grievance Panel after she cited a hallucinated case in her reply brief.⁶ In November 2023, a Colorado attorney received a one-year-and-one-day suspension after using ChatGPT to draft a motion containing fictitious citations, then lying to the court about the source.⁷ In July 2025, three attorneys at a major law firm were disqualified from a civil rights matter after submitting ChatGPT-hallucinated citations, and the district judge warned that monetary sanctions alone were “insufficient deterrents” for AI misuse.⁸ That same month, a Florida federal court dismissed four cases without prejudice and ordered the attorney to pay defendants approximately \$86,000 in fees after he submitted AI-hallucinated citations across eight related cases even after being placed on notice—then the largest monetary sanction imposed for AI citation misuse.⁹ That September, California’s first published appellate opinion on the subject imposed \$10,000 in sanctions and noted that twenty-one of twenty-three case quotations in the appellant’s opening brief were fabricated.¹⁰ That October, a New York state court sanctioned a lawyer who submitted additional AI-hallucinated citations in the very brief opposing the sanctions motion that resulted from his

¹ *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443, 450–51 (S.D.N.Y. 2023).

² *See id.* at 456.

³ *See id.* at 450, 456.

⁴ *Id.* at 452–53, 456.

⁵ Benjamin Weiser, *Here’s What Happens When Your Lawyer Uses ChatGPT*, N.Y. TIMES (May 27, 2023), <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>; *see Mata*, 678 F. Supp. 3d at 466.

⁶ *See Park v. Kim*, 91 F.4th 610, 615–16 (2d Cir. 2024) (holding that Rule 11 requires attorneys to “read, and thereby confirm the existence and validity of, the legal authorities on which they rely”).

⁷ *See People v. Crabill*, No. 23PDJ067, 2023 WL 8111898, at *1 (Off. Presiding Disciplinary J. Sup. Ct. Colo. Nov. 22, 2023) (holding that the attorney violated Colorado Rules of Professional Conduct 1.1, 1.3, 3.3(a)(1), and 8.4(c)).

⁸ *See Johnson v. Dunn*, 792 F. Supp. 3d 1241, 1246–47 (N.D. Ala. 2025) (disqualifying three attorneys and referring them to state disciplinary authorities after they submitted five ChatGPT-hallucinated citations).

⁹ *See ByoPlanet Int’l, LLC v. Johansson*, 792 F. Supp. 3d 1341, 1347, 1354–55, 1357–58 (S.D. Fla. 2025) (finding bad faith where attorney continued submitting ChatGPT-hallucinated citations across eight cases after being placed on notice on April 25, 2025, including in response to the court’s own show cause order; dismissing four federal cases without prejudice and ordering payment of defendants’ attorneys’ fees; and referring attorney to the Florida Bar); *The \$86,000 AI Sanction: ByoPlanet v. Johansson Explained*, NEXLAW: NEXLAW-KNOWLEDGE-CENTER (Sep. 2026), <https://www.nexlaw.ai/blog/byoplanet-ai-hallucination-sanctions/> [<https://perma.cc/MP2U-Y2AY>].

¹⁰ *See Noland v. Land of the Free, L.P.*, 336 Cal. Rptr. 3d 897, 905, 911, 915 (Ct. App. 2025) (imposing \$10,000 in sanctions and notifying the State Bar of California after finding that various case quotations and incorrect citations in the attorney’s briefs were fabricated through multiple AI tools including ChatGPT, Claude, Gemini, and Grok).

prior hallucinated citations.¹¹ In March 2026, the Sixth Circuit in *Whiting v. City of Athens* found over two dozen fake citations and misrepresentations of fact across three consolidated appeals and ordered the responsible attorneys to pay the appellees' full appellate attorneys' fees, double costs, and \$15,000 each in punitive sanctions, observing that "smaller fines have plainly been inadequate" in the face of a "continuous stream" of similar cases.¹² These cases represent a fraction of a documented global phenomenon, with a continuously updated database now cataloging more than 2,000 judicial decisions involving AI-generated hallucinations across at least six countries.¹³

Each incident detailed above generated sanctions or disciplinary referrals. Each was framed, at least initially, as a failure of reasonable inquiry. Courts reached for Federal Rule of Civil Procedure 11's objective reasonableness standard,¹⁴ for 28 U.S.C. § 1927,¹⁵ or for inherent sanctions authority.¹⁶ The American Bar Association's first formal ethics opinion on generative AI told lawyers they can use these tools but must verify their outputs.¹⁷ State bar opinions followed the same logic.¹⁸ The message was consistent: this is a competence problem. Lawyers must be more diligent.

This Essay argues that framing AI citation errors as a negligence problem is the wrong diagnosis. Large language models do not forget. They generate probabilistically plausible text without any concern for factual accuracy, and their propensity to produce false citations was publicly known by 2023, with precise legal-query rates documented in empirical research soon after.¹⁹ An attorney who submits AI-generated citations without independent verification does not

¹¹ See *Ader v. Ader*, No. 653917/2024, 2025 WL 2831332, at *1–2, *4 (N.Y. Sup. Ct. Oct. 1, 2025) (imposing joint and several liability for plaintiff's fees and referring counsel to disciplinary authorities in New York and New Jersey after counsel submitted hallucinated citations in summary judgment brief, then submitted four nonexistent citations and seven fabricated quotations in the opposition to the resulting sanctions motion).

¹² See 170 F.4th 455, 458, 466–68 (6th Cir. 2026) (finding both fabricated cases and real cases cited "for propositions of law that they did not even discuss").

¹³ See Damien Charlotin, *AI Hallucination Cases Database*, <https://www.damiencharlotin.com/hallucinations/> [<https://perma.cc/89FW-25TZ>] (last visited Sep. 7, 2026) (tracking only cases where courts explicitly found reliance on AI-hallucinated content, and thus representing a conservative floor of documented incidents).

¹⁴ See, e.g., *Park v. Kim*, 91 F.4th 610, 614–15 (2d Cir. 2024); see also Fed. R. Civ. P. 11(b) (requiring any attorney signing a pleading to certify that the "legal contentions are warranted by existing law" after "an inquiry reasonable under the circumstances").

¹⁵ See, e.g., *ByoPlanet Int'l, LLC v. Johansson*, 792 F. Supp. 3d 1341, 1353–55 (S.D. Fla. 2025); see also 28 U.S.C. § 1927 (2018) (authorizing courts to require any attorney who "multiplies the proceedings in any case unreasonably and vexatiously" to pay personally the excess costs, expenses, and fees incurred).

¹⁶ See, e.g., *Johnson v. Dunn*, 792 F. Supp. 3d 1241, 1259–60, 1265–66 (N.D. Ala. 2025); see also *Chambers v. NASCO, Inc.*, 501 U.S. 32, 44–50 (1991) (affirming federal courts' inherent power to sanction bad-faith conduct by assessing attorney's fees, and holding that this power survives even where statutory sanctions under Rule 11 or § 1927 are also available).

¹⁷ A.B.A. Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024) (discussing generative AI tools).

¹⁸ See, e.g., Fla. Bar Bd. of Governors, Advisory Op. 24-1 (2024); State Bar of Cal., Standing Comm. on Pro. Resp. & Conduct, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law (Nov. 16, 2023); N.Y.C. Bar Ass'n, Comm. on Pro. Ethics, Formal Op. 2024-5 (2024); Tex. Pro. Ethics Comm., Op. 705 (2025); Pa. Bar Ass'n Comm. on Legal Ethics & Pro. Resp. & Phila. Bar Ass'n Pro. Guidance Comm., Joint Formal Op. 2024-200 (2024).

¹⁹ See, e.g., Weiser, *supra* note 5; *Judge-Specific Requirements: Mandatory Certification Regarding Generative Artificial Intelligence*, U.S. DIST. CT. N. DIST. TEX. (May 30, 2023), <https://www.txnd.uscourts.gov/sites/default/files/documents/CertReStarrJSR.doc> [<https://perma.cc/QEZ4-V42G>] (requiring attorneys to certify either that no portion of a filing was drafted by generative artificial intelligence or that any AI-drafted language was "checked for accuracy, using print reporters or traditional legal databases, by a human being"); Chief Justice John G. Roberts, Jr., *2023 Year-End Report on the Federal Judiciary*, SUP. CT. OF THE U.S. 5–6 (Dec. 31, 2023),

merely fail a diligence standard. They submit a representation to a court while consciously declining to establish any basis for its accuracy. That conduct satisfies the elements of reckless misrepresentation under the Restatement (Second) of Torts, whose scienter vocabulary this Essay borrows without importing the reliance and damages elements of a private fraud claim.²⁰ Where the attorney had constructive notice of the hallucination risk and performed no independent verification, the appropriate legal response is a higher sanction floor, not just a stronger admonition.

The argument does not stand unopposed. Raising the liability standard for AI citation errors could deter adoption of tools that genuinely improve access to justice; for legal aid offices, solo practitioners, and self-represented litigants, AI offers research and drafting capacity at a fraction of conventional cost. Any workable framework must address that concern. This Essay proposes a graduated accountability structure that distinguishes among levels of culpability and preserves those access benefits by sanctioning unverified use in the face of known risk rather than the choice to use AI at all, without allowing willful blindness toward a known category of fabrication to escape the recklessness label.

This Essay proceeds in four Parts. Part I examines the technical structure of AI hallucination, introduces the concept of “structurally predictable error,” and explains why AI citation fabrication is categorically different from human memory error. Part II surveys the current legal framework and identifies its inadequacy. Part III builds the doctrinal case for reckless misrepresentation, drawing on the Restatement (Second) of Torts, the willful blindness doctrine, and the epistemological literature on AI output. Part IV addresses the access-to-justice tension, proposes a three-tier graduated accountability framework with prophylactic disclosure requirements, examines the privilege implications of mandatory disclosure, analyzes whether retrieval-augmented generation changes the recklessness calculus, and surveys comparative international regulatory developments.

I. THE ANATOMY OF AI HALLUCINATION

A. GENERATING LEGAL TEXT

Large language models (LLMs) generate text through a process fundamentally different from search, retrieval, or memory.²¹ They predict, token by token, which sequence of words is statistically probable given their training data and a user’s prompt.²² When a lawyer asks a language model to “find cases on punitive damages in maritime tort law,” the model does not

<https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf> [https://perma.cc/7FUS-ARHB] (warning that “hallucination” had “caused the lawyers using the application to submit briefs with citations to non-existent cases”). For the empirical research documenting precise legal-query hallucination rates, see *infra* notes 24 and 30 and accompanying text.

²⁰ See RESTATEMENT (SECOND) OF TORTS § 526 (A.L.I. 1977).

²¹ This description applies to general-purpose models. Purpose-built legal research tools couple the same generative architecture with a retrieval layer grounded in a licensed document corpus, which reduces but does not eliminate hallucination. See *infra* Part IV.D.

²² See Emily M. Bender, Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* , in FACCT '21: PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 610–11 (2021), <https://dl.acm.org/doi/10.1145/3442188.3445922>; Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation*, 55 ACM COMPUTING SURVS. 248:1, 248:2 (2023).

search a legal database. It produces text that resembles what legal research output looks like, based on patterns in billions of documents processed during training.²³

The term “hallucination” entered the Natural Language Processing literature to describe outputs that are factually incorrect but stylistically fluent.²⁴ Researchers Ji and colleagues, in a 2023 survey of 150 papers, defined hallucination as instances where generated content is “nonsensical or unfaithful to the provided source content.”²⁵ In the legal context, hallucinations take a specific form: the model generates case names, reporter volumes, page numbers, and quotations that read exactly like real citations but correspond to nothing in any official reporter, or real case names attached to propositions those cases do not support—a variant harder to catch because the cited opinion exists.²⁶

The technical cause is architecture. Language models optimize for linguistic plausibility, not factual accuracy.²⁷ The OpenAI GPT-4 Technical Report, published in March 2023, explicitly acknowledged that the model “can suffer from ‘hallucinations’” and identified this as a key safety limitation.²⁸ Researchers Lin, Hilton, and Evans, in a benchmark study using a dataset of 817 questions designed to elicit truthful answers, found that larger models were generally less truthful than smaller ones, possibly because they generated more confident-sounding false claims.²⁹ Nor has the problem proven transitory: successive model generations continue to hallucinate, and courts are still sanctioning fabricated citations in 2026.³⁰ The problem is not incidental to model design. It is structural.

B. MEASURING HALLUCINATION RATES

The frequency of hallucination in legal contexts is not speculative. It is measured. A Stanford and Yale research team tested four leading language models (ChatGPT-4, ChatGPT-3.5, PaLM 2, and Llama 2) on over 800,000 queries across four LLMs (over 200,000 per model) drawn from verifiable factual assertions about randomly selected federal court cases.³¹ They found hallucination rates of between fifty-eight percent to eighty-eight percent across models, with GPT-4 producing false information on fifty-eight percent of queries.³² Varun Magesh and colleagues at

²³ See Bender et al., *supra* note 22, at 610, 616; Matthew Dahl, Varun Magesh, Mirac Suzgun & Daniel E. Ho, *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, 16 J. LEGAL ANALYSIS 64, 66 (2024).

²⁴ See Ji et al., *supra* note 22, at 248:3.

²⁵ *Id.*

²⁶ See Dahl et al., *supra* note 23, at 69–70, 74–76.

²⁷ See Bender et al., *supra* note 23, at 617.

²⁸ OPENAI, GPT-4 TECHNICAL REPORT 1–2 (2023).

²⁹ See Stephanie Lin, Jacob Hilton & Owain Evans, *TruthfulQA: Measuring How Models Mimic Human Falsehoods*, in PROCEEDINGS OF THE 60TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (VOLUME 1: LONG PAPERS) 3214, 3215–16 (Ass’n for Computational Linguistics., 2022), <https://aclanthology.org/2022.acl-long.229.pdf>.

³⁰ See, e.g., *Flycatcher Corp. v. Affable Ave. LLC*, No. 24 Civ. 9429, 2026 WL 306683, at *1–2 (S.D.N.Y. Feb. 5, 2026) (thirteen nonexistent cases and eight real cases that did not contain the attributed quotations; striking submissions and entering default judgment); *Deutsche Bank Nat’l Tr. Co. v. LeTennier*, 250 N.Y.S.3d 260, 265, 268–69 (App. Div. 2026) (“no less than 23 fabricated cases” across five filings; recognizing the case as “the first appellate-level case in New York addressing sanctions for the misuse of GenAI” and sanctioning counsel \$5,000 plus \$2,500); Cynthia Feathers, *Beyond the Mirage: Beware of Generative AI and Hallucinations*, N.Y. STATE BAR ASS’N (June 26, 2026), <https://nysba.org/beyond-the-mirage-beware-of-generative-ai-and-hallucinations/> [<https://perma.cc/28DK-QBJQ>] (surveying the 2025–2026 sanction decisions).

³¹ See Dahl et al., *supra* note 23, at 66, 73.

³² See *id.* at 66, 79–81.

Stanford's RegLab conducted a preregistered empirical study of retrieval-augmented commercial legal research tools that vendors marketed as "hallucination-free."³³ They found hallucination rates between seventeen percent to thirty-three percent, even in tools with retrieval-augmented generation architecture.³⁴

These figures bear directly on professional responsibility. A lawyer who uses a general-purpose language model for legal research faces a greater-than-even chance that any given citation is false.³⁵ A lawyer who uses a specialized legal artificial intelligence (AI) tool faces roughly one-in-six odds.³⁶ The general risk was publicly known by 2023; precise statistics arrived in peer-reviewed research in 2024 and were then covered extensively in bar publications, judicial standing orders, and professional responsibility opinions.³⁷

C. THE CONCEPT OF STRUCTURAL PREDICTABILITY

Human memory error and AI-generated hallucination are not comparable at the mechanical level. When a lawyer incorrectly recalls a case name, the error reflects the genuine fallibility of individual memory applied to the task at hand. The error is idiosyncratic, non-systematic, and not predictable from anything other than the particular facts of that person's cognitive state. It is the kind of error that a reasonable duty of care catches on review.

AI hallucination is categorically different. It does not reflect the failure of any particular instance of cognition. It reflects the architecture of the tool itself. And it evades the review that catches human memory errors: a hallucinated citation arrives fully formatted and confidently quoted, with no facial irregularity for proofreading to notice; only retrieving the cited source exposes it. Emily Bender and colleagues' 2021 paper, which named large language models "stochastic parrots," captured this distinction precisely: these models process text without any reference to meaning and produce outputs that "sound right" without any connection to external

³³ Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. EMPIRICAL LEGAL STUD. 216, 217, 221 (2025).

³⁴ See *id.* at 216–17, 225 fig. 4. For a framework addressing the accountability of AI vendors themselves, see generally Cheng-chi (Kirin) Chang, *Algorithm as Agent: AI Vendor Liability for Title VII Hiring Discrimination After Loper Bright*, 79 STAN. L. REV. ONLINE 50 (2026), <https://review.law.stanford.edu/wp-content/uploads/sites/3/2026/09/Chang-79-Stan.-L.-Rev.-Online-50-1.pdf> [<https://perma.cc/YT2R-GUUG>] (arguing that an AI vendor whose algorithm exercises decisional weight over hiring is the employer's "agent" under Title VII, as distinguished from a mere provider of tools).

³⁵ See Dahl et al., *supra* note 23, at 66.

³⁶ See Magesh et al., *supra* note 33, at 216, 225 fig. 4 (finding that the specialized legal AI research tools made by LexisNexis and Thomson Reuters hallucinated between seventeen percent to thirty-three percent of the time).

³⁷ See A.B.A. Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024); see also *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443, 466 (S.D.N.Y. 2023) (ordering sanctions on attorneys who submitted filing with AI-generated citations to nonexistent cases). *Mata* generated specific guidance from dozens of state bars and coverage in virtually every major U.S. legal publication within weeks of issuance. See, e.g., Sarah Martinson, *ChatGPT Mishap Attys 'Truly Mortified,' Ask for No Sanctions*, LAW360 (June 7, 2023, at 15:50 ET), <https://www.law360.com/articles/1686237/chatgpt-mishap-attys-truly-mortified-ask-for-no-sanctions>; Debra Cassens Weiss, *Lawyers Who 'Doubled Down' and Defended ChatGPT's Fake Cases Must Pay \$5K, Judge Says*, A.B.A. J. (June 26, 2023, at 13:18 CT), <https://www.abajournal.com/web/article/lawyers-who-doubled-down-and-defended-chatgpts-fake-cases-must-pay-5k-judge-says> (reporting the sanctions order and noting contemporaneous coverage by the Volokh Conspiracy, Reuters, and Law360); *supra* note 18; State Bar of Cal., *Artificial Intelligence*, <https://www.calbar.ca.gov/legal-professionals/ethics-compliance-practice-resources/ethics/ethics-technology-resources/artificial-intelligence> [<https://perma.cc/7T26-28QU>] (last visited Aug. 1, 2026) (collecting generative AI ethics guidance from bar authorities nationwide, including Alaska, California, Colorado, the District of Columbia, Mississippi, New Jersey, New Mexico, New York City, Virginia, Washington, and West Virginia).

reality.³⁸ Michael Townsen Hicks and colleagues, in a 2024 paper in *Ethics and Information Technology*, applied Harry Frankfurt’s philosophical concept of “bullshit” (speech produced without concern for truth) to AI outputs, arguing that this quality is more culpable than error, which at least occurs despite genuine concern for truth.³⁹

This Essay introduces the concept of “structurally predictable error” to name this phenomenon. An error is structurally predictable when its occurrence follows necessarily from the architecture of the tool producing it, when its rate is documented and publicly known, and when a reasonable professional in the field either knows or should know its likelihood before using the tool.⁴⁰ AI-generated legal citations meet all three criteria. The user who relies on the output without verification does not exercise poor judgment about a particular matter. They adopt a posture of systematic indifference toward a known category of falsehood.

II. INADEQUACIES OF THE CURRENT LEGAL FRAMEWORK

A. RULE 11 AND OBJECTIVE REASONABLENESS

The dominant response to artificial intelligence (AI) citation hallucinations has deployed Federal Rule of Civil Procedure 11. By signing a pleading, an attorney certifies that the legal contentions are “warranted by existing law” following “an inquiry reasonable under the circumstances.”⁴¹ The Supreme Court confirmed in *Business Guides, Inc. v. Chromatic Communications Enterprises, Inc.* that this standard is objective: sanctions may be imposed not only for bad faith, but for failure to conduct a reasonable prior inquiry.⁴²

Rule 11 is the right starting point, but it reaches only negligence-equivalent conduct. Courts applying it to AI citation errors have treated the failure to verify as analogous to citing a case without reading it. The court in *Mata v. Avianca, Inc.* made this equivalence explicit: the attorneys failed to conduct any verification, which was unreasonable.⁴³ The Second Circuit in *Park v. Kim* focused on the attorney’s failure to “read, and thereby confirm the existence and validity of, the legal authorities on which they rely.”⁴⁴

³⁸ Bender et al., *supra* note 22, at 616–17.

³⁹ See Michael Townsen Hicks, James Humphries & Joe Slater, *ChatGPT Is Bullshit*, 26 ETHICS & INFO. TECH. 37, 37–38 (2024); HARRY G. FRANKFURT, ON BULLSHIT 33–34 (2005).

⁴⁰ The reasonable-professional baseline operates like Rule 11’s objective reasonable-attorney standard and the customary-practice benchmark of legal malpractice: it asks what a competent practitioner in the relevant practice area would know, not what the particular attorney happened to know. Resource constraints bear on sanction magnitude within the first tier, but they do not lower the knowledge baseline. See *infra* Parts IV.A–B. Judges remain divided on how much forewarning the baseline requires. Compare *McCarthy v. U.S. Drug Enf’t Admin.*, 171 F.4th 245, 251 (3d Cir. 2026) (imposing a reprimand and finding it mitigating that the court had “not yet addressed the failure of attorneys to use AI with appropriate guardrails”), with *id.* at 254 (Roth, J., concurring in part and dissenting in part) (“[Courts] need not explicitly forewarn against improper AI use . . .”).

⁴¹ Fed. R. Civ. P. 11(b).

⁴² See 498 U.S. 533, 548–51 (1991) (holding that Rule 11’s certification standard requires that a represented party who signs a paper must conduct a reasonable inquiry before filing, and that sanctions are authorized based on failure to conduct such an inquiry rather than on subjective bad faith).

⁴³ See 678 F. Supp. 3d 443, 456–57, 464 (S.D.N.Y. 2023).

⁴⁴ 91 F.4th 610, 615 (2d Cir. 2024); see also *Coomer v. Lindell*, No. 22-cv-01129-NYW-SBP, 2025 U.S. Dist. LEXIS 128372, at *7–8, *21–23 (D. Colo. July 7, 2025) (noting that “[c]ourts across the country . . . recognize that Rule 11 applies to the use of artificial intelligence,” rejecting counsel’s claim that filing of hallucinated brief was an inadvertent error as contradicted by documentary record, and noting that the same attorneys filed similarly defective briefs in a

The Rule 11 framework captures the omission. It does not capture the character of the representation. When a lawyer submits a citation to a court, they make a positive representation that the cited authority exists and says what they claim. That representation is not merely a failure of process; when the lawyer deliberately chooses not to establish any basis for it, the defect lies in the representation itself. And the representation is made by the lawyer who signs the filing even when another lawyer drafted it, an allocation Part IV.B takes up. The question is whether AI-generated citation errors upgrade that representation from negligent to reckless.

B. MODEL RULE 3.3 AND THE DUTY OF CANDOR

Model Rule of Professional Conduct 3.3 prohibits a lawyer from knowingly making false statements of fact or law to a tribunal.⁴⁵ Comment 3 extends this to statements made “on the lawyer’s own knowledge,” which “may properly be made only when the lawyer knows the assertion is true or believes it to be true on the basis of a reasonably diligent inquiry.”⁴⁶ American Bar Association Formal Opinion 512 interpreted this to require review of all AI-generated citations before submission.⁴⁷

The Rule 3.3 framework focuses on knowledge. It catches the attorney who knows a case does not exist and submits it anyway. It does not reach the attorney who submits an AI-generated citation in conscious ignorance of the tool’s hallucination rate, never asking whether the output is real. “Knowing” under the Model Rules requires subjective awareness.⁴⁸ The attorney who never checks, and who arranges their practice so that they will never know whether their citations exist, occupies a space the Model Rules do not fully address: the Rules trigger on what the lawyer knows, and their ignorance is a deliberate choice that keeps that trigger from firing.

C. STATE BAR GUIDANCE AND PROFESSIONAL COMPETENCE

Forty states and the District of Columbia have adopted the 2012 amendment to Model Rule 1.1 requiring lawyers to keep abreast of changes in technology.⁴⁹ State bars began issuing AI-specific guidance in 2023.⁵⁰ The Florida Bar’s Advisory Opinion 24-1, the California State Bar’s Practical Guidance on Generative AI, and the New York City Bar’s Formal Opinion 2024-5 each concluded that lawyers may use generative AI but must verify their outputs.⁵¹

separate federal action seven days after issuance of the order to show cause, demonstrating a pattern rather than an isolated mistake).

⁴⁵ MODEL RULES OF PRO. CONDUCT r. 3.3(a)(1) (A.B.A. 2023).

⁴⁶ *Id.* cmt. 3.

⁴⁷ See A.B.A. Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024).

⁴⁸ MODEL RULES OF PRO. CONDUCT r. 1.0(f) (A.B.A. 2023) (“‘Knowing’ . . . denotes actual knowledge of the fact in question.”).

⁴⁹ See Bob Ambrogi, *D.C. Adopts Duty of Technology Competence for Lawyers, But Takes a Different Approach*, LAWSTATES (Apr. 24, 2025), <https://www.lawnext.com/2025/04/d-c-adopts-duty-of-technology-competence-for-lawyers-but-takes-a-different-approach.html> [<https://perma.cc/88W4-WY54>]; see also MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (A.B.A. 2023) (“To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice . . .”).

⁵⁰ See *supra* note 18.

⁵¹ See Fla. Bar Bd. of Governors, Advisory Op. 24-1 (2024); State Bar of Cal., Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law (Nov. 16, 2023); N.Y.C. Bar Ass’n, Formal Op. 2024-5, at 4–5 (2024).

These opinions frame the verification duty as a competence obligation. They do not grapple with what it means when a lawyer fails to verify despite having constructive knowledge that AI-generated legal citations carry a substantial and documented risk of fabrication. The competence framing treats non-verification as a skills deficit. This Essay argues it is more accurately characterized as a deliberate choice to operate without concern for the accuracy of representations made to a court.

III. THE CASE FOR RECKLESS MISREPRESENTATION

A. THE ARCHITECTURE OF RECKLESSNESS

Recklessness occupies a conceptual space between negligence and intentional misconduct.⁵² The Restatement (Second) of Torts defines reckless conduct as conduct where the actor “knowing or having reason to know of facts which would lead a reasonable man to realize, not only that his conduct creates an unreasonable risk of physical harm to another, but also that such risk is substantially greater than that which is necessary to make his conduct negligent.”⁵³ Unlike negligence, where the actor simply fails to recognize the risk, recklessness requires awareness of facts from which the risk would be apparent to a reasonable person, combined with a decision to proceed anyway.⁵⁴ It differs from intentional misconduct at the other boundary: the intentional wrongdoer knows the statement is false or means to deceive; the reckless actor proceeds without caring whether it is true.⁵⁵

Reckless misrepresentation occupies the same position in the law of deceit. Restatement Section 526 specifies that a misrepresentation is fraudulent if the maker “does not have the confidence in the accuracy of his representation that he states or implies,” or “knows that he does not have the basis for his representation that he states or implies.”⁵⁶ Section 552, governing information supplied in a professional context, establishes liability when a professional “fails to exercise reasonable care or competence in obtaining or communicating the information.”⁵⁷ That standard defines the negligence floor that courts have so far applied to artificial intelligence (AI) citation errors, and it is the standard this Essay argues is insufficient. The recklessness ceiling is supplied by Section 526(b)–(c)⁵⁸: where the professional not only falls short of due care but submits a representation without any confidence in its accuracy, the conduct exceeds negligence, and the law of deceit supplies the appropriate label.

This Essay does not invoke Section 526 as a cause of action, and no private tort claim is asserted here. The Restatement’s recklessness taxonomy performs a different and more limited function: it supplies a *scienter* vocabulary that courts applying Rule 11, 28 U.S.C. § 1927, and

⁵² This Essay’s framework applies exclusively to attorneys’ duty of candor to tribunals under Model Rule 3.3. The distinct question of judicial officers’ own AI use falls within judicial administration and recusal doctrine, not professional responsibility, and raises no inconsistency with the framework proposed here.

⁵³ RESTATEMENT (SECOND) OF TORTS § 500 (A.L.I. 1965).

⁵⁴ *See id.* cmt. b (“It is reckless for a driver of an automobile intentionally to cross a through highway in defiance of a stop sign if a stream of vehicles is seen to be closely approaching in both directions, but if his failure to stop is due to the fact that he has permitted his attention to be diverted so that he does not know that he is approaching the crossing, he may be merely negligent and not reckless.”).

⁵⁵ *See id.* cmt. f.

⁵⁶ *Id.* § 526.

⁵⁷ *Id.* § 552(1).

⁵⁸ *Id.* § 526.

inherent sanctions authority already consult but have not named. This cross-regime borrowing has Supreme Court precedent. In *Chambers v. NASCO*, the Court authorized inherent sanctions for conduct tantamount to bad faith, a standard it traced directly to the historic power of equity to sanction fraud upon the court.⁵⁹ Sections 526(b) and (c) do not create a new cause of action here; they give a doctrinal name to a scienter category that sanctions regimes already occupy but have articulated only in ad hoc terms.

B. APPLYING THE RECKLESSNESS STANDARD

The attorney who submits AI-generated citations without verification knows or should know several things. They know they used a generative AI tool. They know, or a reasonable professional in their position should know, that these tools produce false citations at rates ranging from seventeen percent to eighty-eight percent, depending on the tool.⁶⁰ They know this risk has been documented in peer-reviewed research, covered in bar journals, addressed in ethics opinions, and sanctioned by federal and state courts in decisions that received wide coverage in the legal press.⁶¹

They submit the citation without verifying it. In that submission, they implicitly represent to the court that the cited case exists and says what they claim.⁶² They do not have “the confidence in the accuracy of [their] representation that [they] state[] or impl[y].”⁶³ They “know that they do not have the basis for their representation that they state or imply.”⁶⁴ These are not the elements of negligence.⁶⁵ They are the elements of reckless misrepresentation under Section 526.

The traditional objection is that recklessness requires conscious awareness of the risk. The attorney did not think “this case may not exist.” They simply did not think about it at all. But this argument proves too much. Courts have long recognized that willful blindness is functionally equivalent to knowledge for purposes of fraud and misrepresentation: the Supreme Court in *Global-Tech Appliances, Inc. v. SEB S.A.* called the doctrine well established.⁶⁶ *United States v. Jewell* holds that deliberate ignorance and positive knowledge are equally culpable.⁶⁷ The attorney who uses ChatGPT for legal citation research in 2025 without asking whether it produces accurate citations has chosen not to know.⁶⁸ The Restatement’s “knows or has reason to know” standard

⁵⁹ See 501 U.S. 32, 44–46 (1991).

⁶⁰ Dahl et al., *supra* note 23, at 79–81; Magesh et al., *supra* note 33, at 216–17, 225 fig. 4.

⁶¹ See, e.g., A.B.A. Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024); Park v. Kim, 91 F.4th 610, 615 (2d Cir. 2024); Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 466 (S.D.N.Y. 2023). The *Mata* sanctions order was covered by virtually every major U.S. legal publication within weeks of issuance. See *supra* note 37.

⁶² See Kohls v. Ellison, No. 24-cv-3754, 2025 LX 108970, at *13–14 (D. Minn. Jan. 10, 2025) (holding that “signing a declaration under penalty of perjury is not a mere formality” and that the court “cannot accept false statements—innocent or not—in an expert’s declaration submitted under penalty of perjury,” extending the same representational obligation to expert witnesses whose AI-generated citations an attorney submits to a tribunal).

⁶³ RESTATEMENT (SECOND) OF TORTS § 526(b) (A.L.I. 1977).

⁶⁴ *Id.* § 526(c) (alterations omitted).

⁶⁵ See *ByoPlanet Int’l, LLC v. Johansson*, 792 F. Supp. 3d 1341, 1355 (S.D. Fla. 2025) (holding that an attorney who continued using AI-generated citations after being explicitly placed on notice had crossed from recklessness into bad faith, as “[a] reasonable attorney, when made aware that his practices were leading to hallucinated cases and quotations, immediately changes course”).

⁶⁶ See 563 U.S. 754, 766 (2011) (“The doctrine of willful blindness is well established in criminal law.”).

⁶⁷ See 532 F.2d 697, 700 (9th Cir. 1976) (en banc) (“[D]eliberate ignorance and positive knowledge are equally culpable.”).

⁶⁸ See *Coomer v. Lindell*, No. 22-cv-01129-NYW-SBP, 2025 U.S. Dist. LEXIS 128372, at *2–3 (D. Colo. July 7, 2025) (noting counsel admitted on the record that he understood his Rule 11 obligations and that use of generative AI “did not excuse” him from those obligations, yet personally performed no citation verification before filing).

encompasses this choice.⁶⁹ The source of the fictitious authority is irrelevant to this analysis: as one court held when sanctioning an attorney who denied using AI, the duty is violated whenever nonexistent case law is submitted to a tribunal, regardless of how the attorney acquired it.⁷⁰

C. EPISTEMOLOGICAL INDIFFERENCE

The recklessness argument gains additional force from epistemological analysis of what AI tools actually do. Luciano Floridi has argued that large language models represent “agency without intelligence,” producing outputs without any internal state corresponding to truth, belief, or concern for accuracy.⁷¹ Hicks and colleagues argue that AI outputs constitute “bullshit” in Frankfurt’s technical sense: speech produced without concern for its truth value, more culpable than error (which occurs despite concern for truth) and different in kind from lying (which requires knowledge of falsity).⁷² Johan Fredrikzon argues in a 2025 study that AI hallucinations evidence “epistemological indifference,” a structural feature of models that produce confident-sounding outputs with no internal grounding in accuracy.⁷³

These analyses converge on a single point relevant to professional responsibility. An attorney who relies on AI output for legal citations does not rely on a tool that might be wrong in the way an associate might be wrong. They rely on a tool that does not know the difference between right and wrong, a tool the architecture thereof renders the distinction meaningless to its operation. An attorney who submits its outputs without independent verification adopts the tool’s epistemological indifference as their own posture toward the court.⁷⁴

D. THE STRUCTURAL PREDICTABILITY TEST

This Essay proposes a three-part test for determining when AI-generated error rises to reckless misrepresentation. First, the error must be structurally predictable: its occurrence must follow from the architecture of the tool rather than from random individual failure. Second, the risk must be publicly documented: the hallucination rate of the tool in question must be available in sources accessible to a reasonable practitioner. Third, the attorney must have had constructive notice: given the coverage of AI hallucination risks in bar publications, judicial opinions, and ethics guidance at

⁶⁹ RESTATEMENT (SECOND) OF TORTS § 500 cmt. a (A.L.I. 1965); see *Amarsingh v. Frontier Airlines, Inc.*, No. 24-1391, 2026 WL 352016, at *7 (10th Cir. Feb. 9, 2026) (holding under Fed. R. App. P. 38’s objective standard that an attorney’s failure to verify AI-generated citations constituted “reckless disregard of the attorney’s duties to the court”).

⁷⁰ See *United States v. Hayes*, 763 F. Supp. 3d 1054, 1067 (E.D. Cal. 2025) (“Citing nonexistent case law or misrepresenting the holdings of a case is making a false statement to a court. It does not matter if [generative AI] told you so.” (quoting Paul W. Grimm, Maura R. Grossman & Daniel G. Brown, *Is Disclosure and Certification of the Use of Generative AI Really Necessary?*, 107 JUDICATURE 68, 75 (2023))).

⁷¹ See Luciano Floridi, *AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models*, 36 PHIL. & TECH. 1, 6 (2023).

⁷² Hicks et al., *supra* note 39, at 37–38; FRANKFURT, *supra* note 39, at 61.

⁷³ Johan Fredrikzon, *Rethinking Error: “Hallucinations” and Epistemological Indifference*, 3 CRITICAL AI, Apr. 1, 2025, at 2–4 (2025), <https://doi.org/10.1215/2834703X-11700255> (arguing that hallucinations are evidence that LLMs are “epistemologically indifferent in that they deal neither in facts nor in fiction” and that, unlike a human speaker who may choose to disregard the truth, “a GPT’s indifference toward truth is architecturally enforced”).

⁷⁴ See *Kohls v. Ellison*, No. 24-cv-3754, 2025 LX 108970, at *10–13 (D. Minn. Jan. 10, 2025) (excluding expert declaration containing AI-hallucinated citations, holding that “when attorneys and experts abdicate their independent judgment and critical thinking skills in favor of ready-made, AI-generated answers, the quality of our legal profession and the Court’s decisional process suffer”).

the time of submission, a reasonable attorney in the relevant practice area would have known the risk.⁷⁵

When all three conditions exist, the attorney's submission of unverified AI citations constitutes a representation made without confidence in its accuracy, satisfying the standard for reckless misrepresentation under Restatement Section 526(b)–(c). This standard does not require proof of subjective awareness. It requires that the risk be objectively foreseeable, documented, and known to the professional community. An attorney who submits AI-generated legal citations in 2025 without verification satisfies this standard.⁷⁶

A short application shows the test at work. An attorney who asks a general-purpose chatbot for authority and files its citations unverified satisfies all three prongs. An attorney who uses a research tool based on retrieval-augmented generation (RAG) and independently retrieves the key authorities fails the test as to the citations they verified; if an unverified one slips through, they present the first-tier case Part IV describes. And an attorney who files new unverified AI citations after a court has already questioned their authorities is past recklessness altogether, in the aggravated tier.

IV. A GRADUATED ACCOUNTABILITY FRAMEWORK

A. ACCESS-TO-JUSTICE CONCERNS

Before endorsing a higher accountability standard, this Essay takes seriously the access-to-justice objection. The Legal Services Corporation's 2022 Justice Gap report found that ninety-two percent of civil legal problems faced by low-income Americans received inadequate or no legal help.⁷⁷ Artificial intelligence (AI) tools offer potential benefits for legal aid organizations and solo practitioners who cannot afford specialized legal research subscriptions.⁷⁸ Colleen Chien and Miriam Kim's 2025 field study of generative AI in legal aid settings found that ninety percent of legal aid professionals reported increased productivity, with AI tools handling document drafting, issue spotting, and client communication.⁷⁹

⁷⁵ Compare *Amarsingh v. Frontier Airlines, Inc.*, No. 24-1391, 2026 WL 352016, at *7 (10th Cir. Feb. 9, 2026) (treating January 2025 as a period of relative novelty and reducing sanctions accordingly), with *Fletcher v. Experian Info. Sols., Inc.*, 168 F.4th 231, 235 (5th Cir. 2026) (declaring as of February 2026 that ignorance of AI hallucination risks is "certainly no longer" an excuse).

⁷⁶ See *Lacey v. State Farm Gen. Ins. Co.*, No. CV 24-5205, 2025 WL 1363069, at *5 n.6 (C.D. Cal. May 5, 2025) (expressly rejecting the argument that sanctions were inappropriate because "the AI hallucinations weren't too far off the mark in their recitations of the substantive law," and characterizing that position as "a pretty weak no-harm, no-foul defense"); *Fletcher*, 168 F.4th at 235 ("If it were ever an excuse to plead ignorance of the risks of using generative AI to draft a brief without verifying its output, it is certainly no longer so.").

⁷⁷ LEGAL SERVS. CORP., THE JUSTICE GAP: MEASURING THE UNMET CIVIL LEGAL NEEDS OF LOW-INCOME AMERICANS 7, 19 (2022), <https://lsc-live.box.com/s/xl2v2uraitobbzrhwtjlg0emp3myz1> [<https://perma.cc/4RZD-VZZ8>].

⁷⁸ See Drew Simshaw, *Access to A.I. Justice: Avoiding an Inequitable Two-Tiered System of Legal Services*, 24 YALE J.L. & TECH. 150, 155–56 (2022); see also Cheng-chi (Kirin) Chang, *AI, Legal Labor, and the Jevons Paradox*, 101 N.Y.U. L. REV. ONLINE 142, 143–44 (2026), <https://nyulawreview.org/wp-content/uploads/2026/08/101-NYU-LRev-Online-142-2.pdf> [<https://perma.cc/UX3T-5HT6>] (arguing that AI efficiency gains will expand rather than contract legal work, in part by activating latent demand among those currently priced out of legal services).

⁷⁹ Colleen V. Chien & Miriam Kim, *Generative AI and Legal Aid: Results from a Field Study and 100 Use Cases to Bridge the Access to Justice Gap*, 57 LOY. L.A. L. REV. 903, 909–10 (2025) (reporting that ninety percent of participating legal aid professionals experienced increased productivity, with AI tools assisting in document drafting, issue spotting, and client communication across 100 documented use cases).

A strict recklessness standard that treats every AI citation error as presumptively blameworthy threatens these access benefits. It may deter risk-averse legal aid lawyers and solo practitioners from using AI tools even where verification steps are reasonable. Drew Simshaw has argued that poorly designed AI regulation could produce a two-tiered legal system in which large firms gain efficiency advantages while under-resourced practitioners fall further behind.⁸⁰ Any accountability framework must account for this risk. The framework below does so by confining monetary sanctions to attorneys who had reason to know the documented risk and did not meaningfully verify, while treating good-faith, partially verified use as ordinary negligence. It also governs licensed attorneys only: it does not raise the standard for self-represented litigants, whose filings courts continue to police under Rule 11's more forgiving pro se application.⁸¹

B. A THREE-TIER ACCOUNTABILITY MODEL

This Essay proposes a graduated framework that calibrates accountability to actual culpability. It has three tiers.

At the first tier, negligence, the attorney uses an AI tool without knowledge of its hallucination rate, performs some verification steps, and submits a brief containing one or more false citations as a result of the tool's failure. The appropriate response is admonition, required correction, and mandatory AI training. Monetary sanctions are disproportionate. After *Mata v. Avianca, Inc.* and the ensuing ethics guidance this tier has narrowed, but it remains open: it plausibly covers the attorney who used a purpose-built tool marketed as database-grounded, who performed real but imperfect verification, or whose error is the mischaracterization variant that an existence check cannot catch.⁸²

At the second tier, recklessness, the attorney uses an AI tool in a legal research context where hallucination risk is documented and publicly known, performs no independent verification of citations, and submits false citations to a court. This tier satisfies the structural predictability test. The appropriate responses are monetary sanctions calibrated to the harm, mandatory disclosure to the opposing party and tribunal, and required Continuing Legal Education in AI competence. The courts in *Park v. Kim*, *Johnson v. Dunn*, and *Noland v. Land of the Free, L.P.* reached this tier

⁸⁰ Simshaw, *supra* note 78, at 174–76.

⁸¹ Courts sanction pro se AI misuse directly under frivolous-conduct rules. See *Julien v. Arthur*, 260 N.Y.S.3d 344, 347 (App. Div. 2026) (“[T]he unverified usage of GenAI to draft an appellate brief containing false information constitutes frivolous conduct warranting the imposition of a sanction, even when the offending party is a pro se litigant.”).

⁸² See, e.g., *Idehen v. Stoute-Phillip*, 86 Misc. 3d 1244(A), 2025 WL 2178395, at *5, *8–9 (N.Y. Civ. Ct. July 29, 2025) (imposing a \$1,000 sanction and referring counsel to the grievance committee when filing cited to real cases for propositions they did not support); *Lexos Media IP, LLC v. Overstock.com, Inc.*, No. 22-2324-JAR, 2026 WL 265581, at *1, *7 (D. Kan. Feb. 2, 2026) (holding that all five signing attorneys violated Rule 11(b) when counsel prompted ChatGPT, “[t]aking the role of a judge,” to write an order denying the motion “with caselaw support” for the desired proposition).

correctly but did not articulate the recklessness basis,⁸³ though the Tenth Circuit has since applied the recklessness label in this context without grounding it in a cross-regime doctrinal framework.⁸⁴

The line between the first and second tiers turns on a single inquiry: whether, at the time of submission, the attorney knew or had reason to know that the tool they used carried a substantial and documented risk of producing false citations. Because that mental state is rarely provable by direct evidence, courts may infer it from three proxy indicators. The first proxy is tool type. An attorney who used a general-purpose language model for legal citation research is presumed to have had constructive notice of the hallucination risk, because peer-reviewed research documenting failure rates exceeding fifty-eight percent for such tools was widely covered in bar publications before 2024; that presumption may be rebutted by evidence that the attorney had no exposure to that coverage. An attorney who used a purpose-built research tool based on retrieval-augmented generation (RAG) begins outside the presumption but loses that position if they performed no spot-verification, because the documented seventeen percent to thirty-three percent false-citation rates in such tools were equally available to a reasonable practitioner.

The second proxy is scope of verification. Verification adequate to negate constructive notice requires confirming, in an authoritative source independent of the AI tool, both that each cited case exists and that it stands for the proposition cited. Checking a case name without retrieving the opinion does not satisfy this standard. The inquiry is *ex ante*: courts should ask what verification protocol the attorney followed before filing, not merely whether a false citation survived it. A protocol meeting this standard that nonetheless fails in a particular instance leaves the attorney in the first tier; the *ex post* existence of the error bears on sanction magnitude, not tier assignment.

The third proxy is coverage. Because the hallucination risk attaches to the tool and the workflow rather than to individual outputs, an attorney who verified some but not all citations retains constructive notice with respect to every citation submitted. Partial verification is therefore relevant to sanction magnitude, not to tier assignment: verified citations eliminated actual risk; unverified ones did not.

Where the submitting attorney did not personally use the AI tool, supervisory liability under Model Rule 5.1(a) governs;⁸⁵ a partner or managing lawyer who, after June 2023, establishes no

⁸³ See *Park v. Kim*, 91 F.4th 610, 615–16 (2d Cir. 2024); *Johnson v. Dunn*, 792 F. Supp. 3d 1241, 1262–65 (N.D. Ala. 2025); *Noland v. Land of the Free, L.P.*, 336 Cal. Rptr.3d 897, 912–13 (Ct. App. 2025); see also *Ader v. Ader*, No. 653917/2024, 2025 WL 2831332, at *3–4 (N.Y. Sup. Ct. Oct. 1, 2025) (imposing joint and several liability for plaintiff’s fees and referring counsel to disciplinary authorities in New York and New Jersey after counsel submitted hallucinated citations in summary judgment brief, then submitted four nonexistent citations and seven fabricated quotations in the opposition to the resulting sanctions motion); *ByoPlanet Int’l, LLC v. Johansson*, 792 F. Supp. 3d 1341, 1354–55 (S.D. Fla. 2025) (finding bad faith where attorney continued submitting ChatGPT-hallucinated citations across eight cases after being placed on notice on April 25, 2025, including in response to the court’s own show cause order; dismissing four federal cases without prejudice and ordering payment of defendants’ attorneys’ fees; and referring attorney to the Florida Bar); *United States v. Hayes*, 763 F. Supp. 3d 1054, 1071–73 (E.D. Cal. 2025) (imposing sanctions and mandating service of the order to D.C. Bar, California Bar, and all judges in the district); *Lacey v. State Farm Gen. Ins. Co.*, No. CV 24-5205, 2025 WL 1363069, at *4 (C.D. Cal. May 5, 2025) (holding that attorneys’ undisclosed AI use combined with resubmission of a brief still containing false citations “demonstrate reckless conduct with the improper purpose of trying to influence [the court’s] analysis,” but imposing sanctions under inherent authority without reference to Restatement § 526).

⁸⁴ See *Amarsingh v. Frontier Airlines, Inc.*, No. 24-1391, 2026 WL 352016, at *7 (10th Cir. Feb. 9, 2026) (holding that an attorney’s failure to verify AI-generated citations constituted “reckless disregard of the attorney’s duties to the court” under Fed. R. App. P. 38’s objective standard).

⁸⁵ See MODEL RULES OF PRO. CONDUCT r. 5.1(a) (A.B.A. 2023); *Cassata v. Michael Macrina Architect, P.C.*, 250 N.Y.S.3d 778, 791–92 (Sup. Ct. 2026).

firm-level measures for verifying subordinate AI output in court filings fails the reasonable assurance requirement and thereby satisfies the constructive notice prong of the Tier 2 analysis.⁸⁶

At the third tier, aggravated recklessness, the attorney satisfies the second tier and compounds the submission with active misrepresentation: either lying about the source when questioned by the court, as in *People v. Crabill*;⁸⁷ or submitting fabricated AI-generated “opinions” as confirmation, as in *Mata*;⁸⁸ or misleading the court as to the source of AI-generated errors when responding to a show-cause order, as in *Fletcher v. Experian Information Solutions, Inc.*;⁸⁹ or fabricating a purportedly corrected version of the defective filing after the court’s order to show cause, as in *Coomer v. Lindell*;⁹⁰ or submitting additional hallucinated citations in direct response to a show-cause order concerning prior hallucinated citations, as in *ByoPlanet International, LLC v. Johansson*;⁹¹ or submitting AI-hallucinated citations in the opposition to the very sanctions motion addressing prior hallucinated citations, as in *Ader v. Ader*.⁹² This tier warrants suspension, referral to disciplinary authorities, and in extreme cases, disgorgement of fees. The Colorado Presiding Disciplinary Judge reached this conclusion in *Crabill*, imposing a one-year-and-one-day suspension on findings that the attorney violated Rules 1.1, 3.3(a)(1), and 8.4(c).⁹³ *Ader v. Ader* illustrates a recurring pattern in this tier: the attorney there submitted fresh AI-hallucinated citations in the brief opposing the sanctions motion arising from prior hallucinations, compounding the original misconduct at each opportunity to correct it.⁹⁴ The graduated framework also carries an upstream signal: by categorizing unverified AI citation use as reckless rather than merely negligent, it provides the legal malpractice insurance market with a doctrinal basis for treating that conduct as a distinct and elevated risk class.

C. MANDATORY DISCLOSURE OBLIGATIONS

The framework requires a fourth element that cuts across all three tiers: mandatory disclosure of AI use in legal filings. When an attorney uses generative AI to draft substantive legal arguments or locate legal authority, they must disclose that use to the tribunal and certify that all AI-generated citations have been independently verified against official sources. Judge Brantley Starr’s 2023

⁸⁶ See *Lacey*, 2025 WL 1363069, at *3–5 (imposing joint and several liability on two law firms where one firm generated AI research without disclosing that fact to co-counsel, and co-counsel incorporated the output without independent verification; holding both firms collectively responsible despite one firm having had no knowledge that AI had been used).

⁸⁷ See No. 23PDJ067, 2023 WL 8111898, at *1 (Off. Presiding Disciplinary J. Sup. Ct. Colo. Nov. 22, 2023).

⁸⁸ See 678 F. Supp. 3d 443, 452 (S.D.N.Y. 2023).

⁸⁹ See 168 F.4th 231, 239–40 (5th Cir. 2026) (imposing enhanced sanctions where counsel “misled, evaded, and violated her duties as an officer of this court” in responding to the court’s show-cause order, noting that forthright acknowledgment would likely have resulted in lesser sanctions).

⁹⁰ See No. 22-cv-01129-NYW-SBP, 2025 WL 1865282, at *3–8 (D. Colo. July 7, 2025) (finding that the “correct” version counsel claimed was inadvertently misfiled bore no corroboration in the documentary record and contained notes reflecting edits made after the order to show cause was issued, and that the same attorneys filed similarly defective briefs in a separate federal court seven days later).

⁹¹ See 792 F. Supp. 3d 1341, 1354–56 (S.D. Fla. 2025) (holding that attorney’s submission of fabricated quotations in response to both a motion to dismiss and an order to show cause regarding his hallucinated citations constituted “crystal-clear evidence that Paul attempted to perpetrate a fraud on the Court and his adversaries,” and dismissing four cases under Rule 11 and the court’s inherent authority).

⁹² See No. 653917/2024, 2025 WL 2831332, at *4–5 (N.Y. Sup. Ct. Oct. 1, 2025).

⁹³ See *People v. Crabill*, No. 23PDJ067, 2023 WL 8111898, at *1 (Off. Presiding Disciplinary J. Sup. Ct. Colo. Nov. 22, 2023).

⁹⁴ 2025 WL 2831332, at *4–5.

standing order in the Northern District of Texas pioneered this model,⁹⁵ and multiple federal districts have since adopted similar requirements.⁹⁶

Mandatory disclosure performs two functions. It creates a record of AI use that enables courts to apply the graduated framework above. It also shifts the calculus for the attorney: an attorney who discloses and certifies verification has satisfied their obligation; one who fails to disclose or falsely certifies has crossed into the third tier.⁹⁷ The disclosure requirement converts the choice to use AI into an acknowledged and accountable act. Rulemaking has begun to diverge on this element: New York's Part 161, effective June 2026, permits AI in court submissions without requiring disclosure, while still obligating attorneys to independently ensure that filings contain no fabricated authority. Part 161 thus confirms the verification floor this framework builds on; the notice function it leaves to individual judges' part rules is precisely the work the disclosure element performs here.⁹⁸

A concern arises that mandatory disclosure might implicate attorney-client privilege or the work product doctrine, on the theory that revealing which AI tools were used could expose confidential client information embedded in research prompts or attorney mental impressions.⁹⁹ Recent case law substantially narrows this concern. In February 2026, Judge Rakoff of the Southern District of New York ruled in *United States v. Heppner* that documents generated using a consumer-level AI tool were not protected by attorney-client privilege or the work product doctrine on three independent grounds: the AI was not an attorney capable of forming an attorney-client relationship; the tool's terms of service permitted third-party data collection, destroying confidentiality; and the materials were not prepared at counsel's direction.¹⁰⁰ Consumer-grade AI tools thus already forfeit privilege protection through their own contractual terms, before any court-imposed disclosure obligation arises.

⁹⁵ U.S. DIST. CT. N. DIST. TEX., *supra* note 19.

⁹⁶ See, e.g., *Fletcher v. Experian Info. Sols., Inc.*, 168 F.4th 231, 233–34 (5th Cir. 2026) (describing the Fifth Circuit's formal study of a proposed circuit-wide AI disclosure rule and its decision to decline adoption on the ground that existing rules were already sufficient); Shannon Capone Kirk, Emily A. Cobb & Amy Jane Longo, *Judges Guide Attorneys on AI Pitfalls with Standing Orders*, ROPES & GRAY LLP (Aug. 2, 2023), <https://www.ropesgray.com/en/insights/alerts/2023/08/judges-guide-attorneys-on-ai-pitfalls-with-standing-orders> [<https://perma.cc/54AE-GT7K>]; Andrew M. Perlman, *The Legal Ethics of Generative AI*, 57 SUFFOLK U. L. REV. 345, 355–56 (2024).

⁹⁷ *Lacey v. State Farm Gen. Ins. Co.*, No. CV 24-5205, 2025 WL 1363069, at *3–4 (C.D. Cal. May 5, 2025) (identifying the “initial undisclosed use of AI” as an independent and primary ground for holding conduct tantamount to bad faith, separate from the failure to verify, and noting that the attorneys had “adequate opportunities — before and after their error had been brought to their attention — to stop this from happening”); see also *Cassata v. Michael Macrina Architect, P.C.*, 250 N.Y.S.3d 778, 791–92, 797 (Sup. Ct. 2026) (supervising attorney and firm violated Rule 5.1 where a junior associate filed unverified AI-generated citations); *Matter of Zareh*, 250 N.Y.S.3d 534, 536, 538 (App. Div. 2026) (reciprocal public censure; counsel “conceded that he had failed in adequately supervising the drafting attorney”); *Deutsche Bank Nat'l Tr. Co. v. LeTennier*, 250 N.Y.S.3d 260, 268–69 (App. Div. 2026) (“As with the work from a paralegal, intern or another attorney, the use of GenAI in no way abrogates an attorney's or litigant's obligation to fact check and cite check every document filed with a court.”).

⁹⁸ N.Y. COMP. CODES R. & REGS. tit. 22, § 161 (2026) (permitting the use of AI tools in court submissions without mandatory disclosure; the accompanying model rule requires attorneys to “carefully review” each submission and “independently ensure” that it contains no “fabricated or fictitious cases, statutes, or other material,” with individual judges retaining discretion to impose stricter part rules).

⁹⁹ See MODEL RULES OF PRO. CONDUCT r. 1.6 (A.B.A. 2023) (imposing duty of confidentiality over “information relating to the representation of a client”); Fed. R. Civ. P. 26(b)(3) (protecting from discovery documents “prepared in anticipation of litigation”). A mandatory AI-use disclosure requirement could, in principle, implicate both provisions to the extent that research prompts embed confidential client information or attorney mental impressions.

¹⁰⁰ 820 F. Supp. 3d 292, 296–97 (S.D.N.Y. 2026).

The disclosure obligation this Essay proposes does not require revealing attorney strategy or client confidences. It requires disclosing a method of production: that an AI tool was used to generate the citations at issue. Courts have long held that privilege protects communications and mental impressions, not the fact that a particular research technique was employed. An attorney who uses an enterprise-grade tool with contractual confidentiality protections may retain greater work product protection for underlying research; the disclosure of AI use in the final filing is governed by the tribunal's certification requirements rather than by privilege doctrine.¹⁰¹

D. TECHNOLOGICAL SAFEGUARDS AND RAG MODELS

This Essay acknowledges that hallucination rates and mitigation techniques continue to change. Retrieval-augmented generation (RAG) grounds a model's output in documents retrieved from a defined corpus (in legal tools, a licensed database of actual case law) before generation.¹⁰² Because the final step remains generative, RAG substantially reduces but does not eliminate hallucination rates in legal contexts. The Magesh study found hallucination rates between seventeen percent to thirty-three percent even in RAG-based legal research tools.¹⁰³ Sebastian Farquhar and colleagues published a hallucination detection method in *Nature* in 2024 using semantic entropy to identify confabulations before they reach users.¹⁰⁴ Jonathan Choi and colleagues' 2024 randomized trial found that AI assistance significantly increased legal analysis speed without sacrificing quality: the tools carry genuine benefits worth preserving.¹⁰⁵ Richard Susskind has argued for more than a decade that technology-resistant lawyers ultimately harm clients.¹⁰⁶

The framework should therefore be technology-sensitive.¹⁰⁷ An attorney who uses a purpose-built, RAG-based legal research tool with documented low hallucination rates and performs spot-

¹⁰¹ See *Upjohn Co. v. United States*, 449 U.S. 383, 395, 400–01 (1981) (holding that attorney-client privilege protects only communications, not the underlying facts communicated, and that work product protection under Rule 26(b)(3) extends to materials revealing an attorney's mental processes, conclusions, and legal theories, but not to the bare fact that a particular research method was employed); see also *Heppner*, 820 F. Supp. 3d at 297 (noting the outcome might have been different had counsel directed the client's AI use as part of the representation, suggesting that enterprise tools operated under attorney supervision may retain greater work product protection). Requiring disclosure that AI was used to generate citations reveals a method of production, not a privileged communication.

¹⁰² Patrick Lewis et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, in PROCEEDINGS OF THE 34TH INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 9459, 9459–60 (2020). For empirical support, see Faiz Surani & Daniel E. Ho, *AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries*, STAN. HAI (May 23, 2024), <https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries> [<https://perma.cc/T44J-X7UM>] (finding that RAG-based legal research tools hallucinate markedly less than GPT-4 but still err in one out of six or more legal queries and concluding that RAG is “[n]ot a [p]anacea”); Magesh et al., *supra* note 33, at 216.

¹⁰³ Magesh et al., *supra* note 33, at 216.

¹⁰⁴ Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn & Yarin Gal, *Detecting Hallucinations in Large Language Models Using Semantic Entropy*, 630 NATURE 625, 626 (2024) (introducing a semantic entropy method that detects “confabulations” by measuring disagreement across independently sampled outputs, and finding that the method identifies hallucinations before they reach end users).

¹⁰⁵ Jonathan H. Choi, Amy B. Monahan & Daniel Schwarcz, *Lawyering in the Age of Artificial Intelligence*, 109 MINN. L. REV. 147, 182 (2024) (finding in a randomized controlled trial that GPT-4 significantly increased attorney speed on legal analysis tasks without a statistically significant reduction in output quality).

¹⁰⁶ See RICHARD SUSSKIND, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE 3 (3d ed. 2023) (“To meet the needs of clients, we will need . . . [to] reinvent the way in which legal services are delivered.”).

¹⁰⁷ See *Kohls v. Ellison*, No. 24-cv-3754, 2025 LX 108970, at *10–11 (explaining that “AI, in many ways, has the potential to revolutionize legal practice for the better,” but excluding expert testimony where AI was used without

verification of cited cases occupies a different risk profile from one who uses a general-purpose chatbot with no verification. First-tier negligence is more likely where the tool is purpose-built and verification was attempted; second-tier recklessness is more likely where the tool is general-purpose and no verification occurred.

The critical question is whether use of a RAG-based tool automatically downgrades conduct from second-tier recklessness to first-tier negligence. The answer is no. The structural predictability prong of the three-part test focuses on whether the hallucination risk was publicly documented and known to a reasonable practitioner, not on the underlying architecture. A seventeen percent to thirty-three percent false-citation rate, as measured in the Magesh study across RAG-based legal research tools, remains a substantial and publicly known risk.¹⁰⁸ Architecture is one variable in the recklessness calculus, not a dispositive one; what governs is the combination of documented risk level and the verification steps actually taken. An attorney who uses a RAG-based tool and performs no independent verification at all remains in the second tier.¹⁰⁹

E. COMPARATIVE INTERNATIONAL DEVELOPMENTS

The United States is not alone in confronting AI-generated citation errors, and international comparison serves a specific purpose: it shows the two levers other systems have pulled, structural regulation of the tools and enforcement of existing professional duties, and thereby isolates what neither supplies, a culpability doctrine calibrated to the lawyer's representational choice. The European Union (EU) has adopted the most structurally comprehensive response. The EU Artificial Intelligence Act, Regulation (EU) 2024/1689, classifies as high-risk any "AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts," subject to mandatory human oversight, technical documentation, and EU database registration, with compliance obligations becoming mandatory on August 2, 2026.¹¹⁰ General-purpose AI models such as ChatGPT and Claude became subject to EU transparency obligations on August 2, 2025, and the EU approach places obligations on AI providers and deployers rather than on individual legal practitioners.¹¹¹

independent verification, illustrating that an appropriate framework must distinguish between beneficial AI use and abdication of professional judgment).

¹⁰⁸ See Magesh et al., *supra* note 33, at 216; Lewis et al., *supra* note 102 (describing the RAG architecture in which a language model queries an external database before generating text).

¹⁰⁹ See *ByoPlanet Int'l, LLC v. Johansson*, 792 F. Supp. 3d 1341, 1355 (S.D. Fla. 2025) ("A reasonable attorney does not blindly rely on AI to generate filings. A reasonable attorney, when made aware that his practices were leading to hallucinated cases and quotations, immediately changes course."); Farquhar et al., *supra* note 104 (noting that hallucination probability varies by query type and retrieval quality rather than by architecture alone). Architecture is one variable in the recklessness calculus, not a dispositive one. An attorney who uses an RAG-based tool and performs no independent verification at all remains in the second tier regardless of the tool's claimed accuracy; what governs is the combination of documented risk level and the verification steps actually taken.

¹¹⁰ Council Regulation 2024/1689 of 13 June 2024, Artificial Intelligence Act, annex III, 2024 O.J. (L 12.7) point 8(a) (classifying judicial-assistance AI as high-risk subject to mandatory human oversight under Article 14, technical documentation under Article 11, and EU database registration under Article 49; full compliance required August 2, 2026).

¹¹¹ *Id.* art. 53 (requiring GPAI model providers to publish training-data summaries and comply with copyright law; applicable to models deployed within the EU, effective August 2, 2025); see generally Cheng-chi (Karin) Chang, *The*

The United Kingdom has followed a more principles-based trajectory. The Law Society's 2024 Generative AI guidance confirms that professional duties apply fully regardless of whether AI tools were used and that lawyers must verify all court submissions.¹¹² The Solicitors Regulation Authority has instructed compliance officers to take formal responsibility when new technology is introduced.¹¹³ In June 2025, the Divisional Court of the King's Bench Division addressed AI citation failures directly in *Ayinde v. London Borough of Haringey* and companion case *Al-Haroun v. Qatar National Bank*, referring solicitors to the Solicitors Regulatory Authority (SRA) and barristers to the Bar Standards Board, warning that future cases of AI misuse may constitute contempt of court, and characterizing the filing of fabricated citations as "wholly improper."¹¹⁴ Neither the EU's structural regulation nor the UK's principles-based approach has yet developed a recklessness-equivalent doctrine imposing graduated sanctions calibrated to the nature of the representational choice. The framework this Essay proposes fills that doctrinal gap.

CONCLUSION

Steven Schwartz stood before Judge Castel having submitted six cases that existed nowhere and, when questioned, fabricated opinions to defend them. *Mata v. Avianca, Inc.* itself thus involved more than an unverified first filing. But the broader doctrinal and professional-responsibility response, in *Mata* and the cases that followed, has framed AI citation errors as failures of reasonable inquiry, diligence, or competence. This Essay has argued that framing is wrong. When lawyers use tools known to fabricate authority at documented rates, submit the output without verification, and thereby represent to a court that nonexistent authority says what they claim, the conduct fits the elements of reckless misrepresentation under the Restatement (Second) of Torts, not merely a failure of diligence. The attorney's ignorance of the hallucination is not exculpatory. It is the point.

The reconceptualization this Essay proposes rests on three interlocking arguments. AI hallucination is structurally predictable: the error follows from the tool's architecture, its rate is

First Global AI Treaty: Analyzing the Framework Convention on Artificial Intelligence and the EU AI Act, 2024 U. ILL. L. REV. ONLINE 86 (analyzing the Council of Europe's Framework Convention on AI alongside the EU AI Act and assessing their enforcement and implementation challenges).

¹¹² *Generative AI – the Essentials*, THE L. SOC'Y (Oct. 1, 2025), <https://www.lawsociety.org.uk/topics/ai-and-lawtech/generative-ai-the-essentials> [<https://perma.cc/92QD-NPBP>] (confirming that professional duties "apply to work carried out by the solicitor . . . regardless of whether AI or other technologies were used" and that lawyers "must take positive steps to make sure that any information or documents that you submit to the court are accurate and from genuine and verifiable sources," and characterizing AI-related professional breaches as "extremely serious" with courts empowered to consider the full range of sanctions including contempt).

¹¹³ *Compliance Tips for Solicitors Regarding the Use of AI and Technology*, SOLICS. REGUL. AUTH. (Feb. 9, 2026), <https://www.sra.org.uk/solicitors/resources/innovate/compliance-tips-for-solicitors/> [<https://perma.cc/9V8B-6LX9>] (stating that the Solicitors Regulation Authority (SRA) expects compliance officers to "be responsible for regulatory compliance when new technology is introduced," and confirming that SRA Principles, including duties to uphold the administration of justice and act with integrity, apply to all AI-assisted work). Neither the EU's structural regulation nor the UK's principles-based approach has yet developed a recklessness-equivalent doctrine imposing graduated sanctions calibrated to the nature of the representational choice.

¹¹⁴ *Ayinde v. London Borough of Haringey & Al-Haroun v. Qatar Nat'l Bank QPSC* [2025] EWHC (Admin) 1383 [9], [48], [2025] 1 WLR 5147 (Eng.) (referring solicitors to the Solicitors Regulation Authority and barristers to the Bar Standards Board after fictitious citations appeared in court filings in two separate matters; warning that future cases of AI misuse may constitute contempt of court; stating that "[t]here are serious implications for the administration of justice and public confidence in the justice system if artificial intelligence is misused," and characterizing the filing of fabricated citations as "wholly improper").

publicly documented, and a reasonable practitioner either knows or should know it before using the tool. That structural predictability, combined with an attorney's decision to submit unverified output, satisfies the elements of reckless misrepresentation under Restatement Section 526(b)–(c), supplying the doctrinal name that courts have reached for but not found. And the graduated three-tier framework ensures that the higher accountability standard does not penalize good-faith efforts to verify AI-generated content or foreclose the access-to-justice benefits that these tools can offer. What it does not tolerate is systematic indifference to a known and documented category of fabrication. Courts on both sides of the Atlantic already sense this distinction, and some have begun to name it. This Essay gives that instinct a doctrinal home.