A Typology of Typos:
Detecting Hallucinated Legal Citations Is Harder Than You Think

Benjamin L.W. Sobel,* Brandon Liu,† Sunoo Park‡

Abstract

As lawyers employ error-prone AI, courts face a barrage of filings that cite nonexistent cases as if they were authoritative sources. But the scale of the problem remains unknown. This Article discusses original empirical research designed to measure the rate of hallucinated citations in legal filings on an original dataset of millions of filings spanning 2015–2025. Perhaps surprisingly, we detect *no* spike in "invalid" case citations. Why might this be? First, humans are bad at citations too: across *all* years in our sample—including those before the proliferation of generative AI—we find that roughly 6% of citations are defective. Second, distinguishing hallucinations from human error at scale is remarkably challenging.

Our results raise novel theoretical and doctrinal questions. Courts and scholars decry "fake" citations—but what makes a citation "fake" in the first place? What distinguishes AI hallucinations from defective citations by human authors? To address these questions, we present a novel typology. Successful citations must be both (1) "referentially valid," in that they effectively refer to some existing authority; and (2) "propositionally valid," in that they characterize that authority defensibly. Prior scholarship focuses on propositional validity and elides referential validity, because it overlooks that technically imperfect citations can and often do function in practice.

Legal institutions may need to develop systems to screen for citation validity, and judges are beginning to fashion standards for policing invalid citations. We identify an emerging split: jurists disagree on whether to treat text differently because it was generated by AI. We propose that there is no distinction in form between hallucinations and other citation errors. Thus, courts should not police citation errors based on their provenance. Finally, we reflect on longstanding aspects of our legal infrastructure that have made caselaw so difficult to find and to cite, and we discuss structural solutions.

---

* Assistant Professor of Law, University of Wisconsin Law School. We are grateful to Sam de Alfaro for excellent research assistance, and Andrés Fábrega, James Grimmelmann, Thibaut Horel, Michael Lissner, Winston Luo, Moira Marquis, and the Free Law Project for helpful discussions, feedback, and resources.

† J.D. Candidate, NYU School of Law; Ph.D. Student, Department of Information Science, Cornell Tech.

‡ Assistant Professor, NYU Courant School (Computer Science) & NYU School of Law. This research was partially supported by NYU IT High Performance Computing resources, services, and staff, by an unrestricted gift from Google, and by the Institute of Information & Communications Technology Planning & Evaluation (IITP) with a grant funded by the Ministry of Science and ICT (MSIT) of the Republic of Korea in connection with the Global AI Frontier Lab International Collaborative Research.

Table of Contents

I. THE PRACTICAL AND THEORETICAL CHALLENGES OF VALIDATING CITATIONS 6

A. PRACTICAL CHALLENGES (AND TECHNICAL DETAILS) 9
1. The Data 9
2. The Tool 11
a) Document Processing and Citation Extraction 11
b) Building a Citation Database 12
(1) What's in the Database? 12
(2) What's Not in the Database? 13
c) Citation Validation 13
d) Benchmarking and Performance 14
3. The Results 15
a) Real Citations That Failed To Validate 16
b) Truly Invalid Citations 16
c) Valid Citations May Be Formally Defective 17
B. THEORETICAL CHALLENGES 18
1. Defining "Citation" 18
2. Defining Citation "Validity" 19
a) Referential Validity 20
(1) Existing Studies Tend to Elide Referential Validity 21
(2) Referential Validity Remains an Unavoidable Problem 24
b) Propositional Validity 25
C. WHAT, EXACTLY, IS A "HALLUCINATION"? 27
1. The Formal Content of a "Hallucination" Is Hard to Specify 28
2. Hallucinated Case Citations Cannot Be Identified from Text Alone 31
a) "Chelates v. Citrate" and Other Stories 32
b) Programmatic Hallucination Detection Is Likely Impossible 34

II. THE DOCTRINAL STAKES OF CITATION VALIDITY 37

A. PROFESSIONAL CONDUCT: COMPETENCE, SUPERVISION, AND CANDOR 38
B. FEDERAL RULES OF CIVIL PROCEDURE: RULE 11(B) 39
C. THE JURISPRUDENCE OF TYPOS 42
D. DOES IT MATTER WHETHER AI INTRODUCED THE ERROR? 44
E. SHOULD IT MATTER WHETHER AI INTRODUCED THE ERROR? 48
F. THE STAKES OF ERRONEOUS PRECEDENT IN COMMON LAW 49

III. REFLECTING ON INVALID CITATIONS AND PRACTICAL CONSIDERATIONS 52

A. ACCESS TO JUSTICE 52
B. STREAMLINING CITATION CONVENTIONS 55
C. MITIGATION STRATEGIES 56
D. WHAT'S THE POINT OF CITATIONS? DO THEY MATTER IN PRACTICE? 58
E. WHO BEARS THE BURDEN? WHO SHOULD BEAR THE BURDEN? 59

CONCLUSION 60

Introduction

In late 2024, it seemed as though the problem of "hallucinated" citations in legal filings might be reaching a peak. The legal community was on notice of the extravagant risks of incautious AI-assisted drafting: eighteen months before, observers had gawked at the first high-profile sanctioning of attorneys who filed documents that "cited . . . fake cases" hallucinated by generative AI.[1] The general public—even as it embraced greater and greater reliance on AI tools—was gaining awareness of AI chatbots' shortcomings, including their propensity to output authoritative-looking quotes and citations that pointed to nonexistent sources. Legal informatics services were rolling out proprietary AI tools that they advertised as "hallucination free."[2] While hallucinated caselaw finding its way into legal filings was clearly a serious problem, at least to these optimistic and informed researchers, it seemed like a problem that might wane as a better-informed, newly cautious cohort of practitioners began to make use of improved tools.[3]

---

[1] Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 449 (S.D.N.Y. 2023). The court noted that the attorneys' misbehavior was not simply filing cases with hallucinated citations: "[I]f the matter had ended with Respondents coming clean about their actions shortly after they received the defendant's . . . brief questioning the existence of the cases, or after they reviewed the Court's Orders . . . requiring production of the cases, the record now would look quite different. Instead, the individual Respondents doubled down and did not begin to dribble out the truth until . . . after the Court issued an Order to Show Cause why one of the individual Respondents ought not be sanctioned." *Id.*

[2] *See* Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. EMPIRICAL LEGAL STUD. 216, 234 n.2 (2025). (collecting reliability claims by legal AI companies).

[3] That said, the perspective that hallucinations may be inherent to generative AI in the foreseeable future had already emerged, and we are open to that possibility. *See, e.g.*, Ziwei Xu, Sanjay Jain & Mohan Kankanhalli, *Hallucination Is Inevitable: An Innate Limitation of Large Language Models* (Feb. 13, 2025) (unpublished manuscript), https://arxiv.org/pdf/2401.11817. *But see* OpenAI, *Why Large Language Models Hallucinate* (Sep. 5, 2025), https://openai.com/index/why-language-models-hallucinate. Our hope was not so much that hallucinations would go away entirely, but that they would decrease

So, over a casual dinner that December, some of the authors of this Article resolved to strike while the iron was hot. We spun out an idea that, at the time, seemed simple: Why not collect a large sample of legal filings—spanning the period before and after widespread public adoption of generative-AI chatbots—and then analyze those filings at scale to detect whether the incidence of "fake" case citations changed over the relevant period? Case citations are standardized enough, we figured, and the universe of potential citations circumscribed enough, that neither the programmatic recognition of citations nor the automated verification of their accuracy should prove all that troublesome. Such a study could provide a systematic view of generative AI's impact on citations in legal filings, thus far documented predominantly through anecdotal evidence of the incidents that happen to catch the attention of a court, opposing counsel, or the news media.

The following year and a half revealed that all of our suppositions were mistaken.[4] First, hallucinated citations were nowhere close to peaking—at least, not as measured by gravity or ignominy. Through 2025 and 2026, hallucinated citations would afflict the docketed work product of at least two federal judges[5] and the "elite Wall Street law firm" Sullivan & Cromwell,[6] not to mention countless hapless attorneys in practice worldwide.[7] Second, detecting hallucinated citations was nothing like the

in prevalence and severity and that cautious practitioners would find better ways to manage them.

[4] *See, e.g.*, Patty Liu, Dominik Stammbach & Peter Henderson, *Who Checks the Citations? Benchmarking Legal Hallucination Detection*, *in* PROCEEDINGS OF THE THIRD CONFERENCE ON LANGUAGE MODELING (forthcoming 2026) (manuscript at 2), https://arxiv.org/pdf/2606.21155 ("A prevailing argument has been that this problem is temporary, that hallucination rates would diminish as models improve, and that high-profile sanctions would induce greater caution. Our findings and recent events challenge these assumptions. Real-world court filings containing hallucinated citations are growing steadily, and the problem is not yet self-correcting.").

[5] Letter from Judge Henry T. Wingate to the Honorable Robert J. Conrad, Jr. (Oct. 21, 2025), https://www.grassley.senate.gov/imo/media/doc/wingate_to_grassley_re_judiciary_use_of_ai.pdf; Letter from Judge Julien X. Neals to the Honorable Robert J. Conrad, Jr. (Oct. 20, 2025), https://www.grassley.senate.gov/imo/media/doc/neals_to_grassley_re_judiciary_use_of_ai.pdf.

[6] Santul Nerkar, *A.I. 'Hallucinations' Created Errors in Court Filing, Top Law Firm Says*, N.Y. TIMES (Apr. 21, 2026), https://www.nytimes.com/2026/04/21/nyregion/sullivan-cromwell-ai-hallucination.html.

[7] Damien Charlotin, *AI Hallucination Cases*, https://www.damiencharlotin.com/hallucinations/ (a collection of nearly two thousand legal cases in which a party relied on hallucinated material).

simple task we surmised it would be. The structure of American courts[8] and their record systems made our primary sources difficult to access and to process. The dominance of proprietary legal-informatics services made a significant portion of citations impossible to verify programmatically. Abstruse citation-formatting conventions and rampant human errors in legal filings made it difficult to measure the impact of generative AI. And perhaps most importantly, basic ontological questions—*What does it mean for a legal citation to be "valid" or "invalid?" What does it mean to "refer to" something in the first place? What makes an error attributable to a familiar human mistake versus an AI hallucination?*—made it difficult to interpret the significance of our results.

We soon intend to publish the technical results of our study in a computer-science venue.[9] While that analysis never pinpointed the silver-bullet statistic we set out to ascertain—"What is the prevalence of citations to non-existent caselaw in legal filings over the last 10 years (before and after the introduction of generative AI)?"—it made a number of contributions to the emerging scholarly literature on AI and law, principal among them the design and open-source implementation of a new citation-verification tool called Hallucite that performs comparably to or better than available alternatives according to a range of performance metrics. But that project also revealed a number of insights about our dysfunctional legal-citation ecosystem, and the very nature of citation as a practice, that warrant elaboration for a legal audience.

This Article presents our original empirical work to detect hallucinated citations and its lessons for the practice of law, legal informatics, and access to justice. Think of it as a scholarly *Lost in La Mancha*[10]—a reflection on the structures that turned a simple empirical goal into an impossibly complicated task, and a discussion of why those structures work against the interests of lawyers, litigants, judges, and the public at large. It proceeds in three Parts, tailored to a legal audience rather than a technical one. Part I presents the problem: why, as an empirical and theoretical matter, is it so hard to determine whether a citation is "hallucinated?" It glosses the technical difficulties our project encountered and the information-access barriers that our legal-informatics economy presented. It then offers a typology of citation failures and presents our

[8] We focus on U.S. federal and state courts.

[9] Brandon Liu et al., Hallucite: New Tools and Benchmarks for Detecting Hallucinated Legal Citations On Real-World Court Filings (July 29, 2026) (unpublished manuscript) (on file with author).

[10] LOST IN LA MANCHA (2002).

conjecture that there is no formal distinction between hallucinations and other citation errors: whether a citation is an AI hallucination or the result of a more familiar human error is probably impossible to ascertain from inspecting the suspect text alone.

Part II identifies a split among courts' and commentators' approaches to policing defective citations. Judges are beginning to fashion standards for sanctioning attorneys who submit invalid citations,[11] and legal institutions may need to develop systems to screen citation validity at scale. But jurists disagree about whether identical text should be treated differently depending on whether it was generated by AI.[12] Because there is no way to programmatically differentiate hallucinations from citations that are erroneous for other reasons, a disciplinary system that seeks to distinguish between citation errors based on their provenance will face serious practical obstacles. Accordingly, we suggest that sanctions and other consequences should depend on the filings that attorneys submit, rather than guesses at, or time-consuming forensic investigations of, the manner in which those filings were prepared.

Part III discusses pragmatic interventions that might help solve the problem and enable higher quality, more accessible legal filings—regardless of who drafted them, and by what means. A brief Conclusion follows.

I. The Practical and Theoretical Challenges of Validating Citations

What is the point of a legal citation? According to the *Bluebook*, which, for better or worse, is the authoritative citation manual in the United States, "The central function of a legal citation is to allow the reader to efficiently locate the cited source."[13] Citations are indispensable to legal practice, adjudication, and scholarship. Indeed, attorneys who cite authority inadequately or improperly risk outcome-determinative consequences and professional sanctions.[14] In a common-law system like those in the United States, citations to precedential caselaw are particularly important; it is hard to imagine an effective legal brief or decision that cites

[11] *See, e.g.*, *AI in State Courts*, NAT'L CTR. FOR STATE CTS., https://www.ncsc.org/resources-courts/ai-state-courts (last visited June 3, 2026).
[12] *See* Part II.D, *infra*.
[13] *Introduction*, THE BLUEBOOK, https://www.legalbluebook.com/introduction (last visited June 3, 2026).
[14] *See* notes 71, 177, *infra*.

no caselaw. It's the *law*, after all; you can't make it up as you go along,[15] and that means that you must[16] cite legal authorities that support your arguments. Furthermore, citations to fake precedents are particularly dangerous in a common-law system, because they distort the foundation upon which future doctrine is based.

But what *is* a legal citation? What is the information that allows a reader to locate a legal source? Our empirical work focuses on caselaw citations, which usually specify: (a) the name of a case, usually that of one or more of the parties (the "caption"); (b) the volume of one or more reporters in which the decision was published; (c) the page number of the cited volume on which the cited material appears; (d) the court that issued the decision; and (e) the date of the decision, as shown in Figure 1. But that is not the end of the story. Here, again, the *Bluebook* is authoritative, this time for worse and not for better. The *Bluebook* imposes "a comically elaborate thicket of random and counterintuitive rules" governing how cases are to be cited in legal documents.[17] It specifies such indispensable details as how party names and reporters are to be abbreviated; the acceptable width of the horizontal line used to indicate a numerical range (try not to use the hyphen, and do not—*do not*—use the em-dash); whether and when certain punctuation marks are to be italicized; and whether one ought to cite the *Vermont Reporter* or West's *Atlantic Reports* when referring to a decision of the Vermont Supreme Court.

[15] *But see, e.g.*, Trump v. United States, 603 U.S. 593 (2024). *Compare, e.g.*, Trump v. Cook, No. 25A312, 2026 WL 1855613, at *9 (U.S. June 29, 2026), *with* Trump v. Slaughter, No. 25-332, 2026 WL 1855612, at *11 (U.S. June 29, 2026). *Compare also, e.g.*, Trump v. CASA, Inc., 606 U.S. 831, 837-39 (2025), *with* Nat'l Institutes of Health v. Am. Pub. Health Ass'n, 145 S. Ct. 2658 (2025); *id.* at 2675 (Jackson, J., concurring in part and dissenting in part) ("This is Calvinball jurisprudence with a twist . . . .").

[16] *But see, e.g.*, Trump v. Wilcox, 145 S. Ct. 1415, 221 L. Ed. 2d 985 (2025) (overruling *Humphrey's Executor v. United States* by fiat, without citing the case); *id.* at 1419 (Kagan, J., dissenting) ("[T]he order allows the President to overrule *Humphrey's* by fiat . . . .").

[17] Adam Liptak, *Yale Finds Error in Legal Stylebook: Harvard Did Not Create It*, N.Y. TIMES (Dec. 7, 2015), https://www.nytimes.com/2015/12/08/us/politics/yale-finds-error-in-legal-stylebook-harvard-did-not-create-it.html; *see also* Richard A. Posner, *The Bluebook Blues*, 120 YALE L.J. 850 (2011).

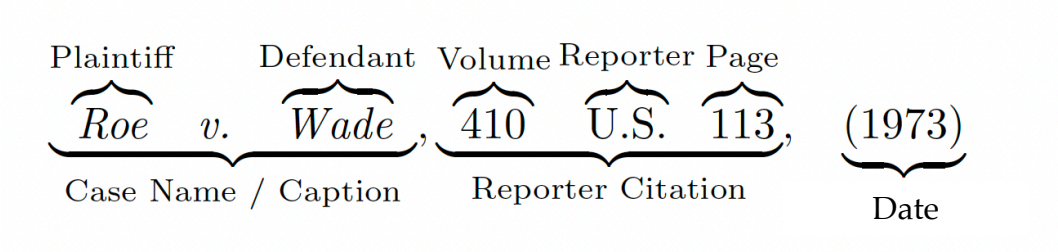


Figure 1: Anatomy of a Caselaw Citation in Bluebook Format[18]

In our opinion—an opinion shared by some eminent jurists[19]—the *Bluebook* exists to provide a busywork screening ritual for aspiring law-review editors and to rent-seek by charging money for access to a system that is not subject to exclusive rights.[20] As the embodiment of "a uniform system of citation,"[21] it has lost the plot completely. Its byzantine mandates have, in all likelihood, precipitated a confusion of tongues: because ascertaining how to cite a case in exact accordance with the *Bluebook*'s many regulations can be very difficult, the manual has probably encouraged the accumulation of small errors and inconsistencies in legal citations that a less convoluted citation system could avoid.

To its credit, however, there is one formula the *Bluebook* hasn't overcomplicated: the volume-reporter-page number format of referring to a case. This information is the crux of a caselaw citation, because it tells the reader the precise page of the precise book on which the cited material is to be found. And here, the *Bluebook*'s abstruse abbreviation guidelines are probably an asset, because they give legal writers short, standardized language for referencing ubiquitous publications. (For example, the second

[18] THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION R. 10 (Columbia L. Rev. Ass'n et al. eds., 22d ed. 2025) [hereinafter BLUEBOOK].

[19] See, e.g., Posner, *supra* note __; Richard A. Posner, *Goodbye to the Bluebook*, 53 U. CHI. L. REV. 1343 (1986).

[20] David Post, *The New (and Much Improved) "Bluebook" Caught in the Copyright Cross-Hairs; Harvard Law School Stands between Us and a Free, Sensible, Open-Access System of Legal Citation.*, WASHINGTON POST.COM, Feb. 11, 2016, http://global.factiva.com/redir/default.aspx?P=sa&an=WPCOM00020160209ec29004pj&cat=a&ep=ASE; *see also* Letter from Peter M. Brody to Carl Malamud and Christopher Jon Sprigman (Dec. 24, 2015), https://law.resource.org/pub/us/code/blue/harvard.response.20151224.pdf (letter from the counsel for the Harvard Law Review Association to the creators of the *BabyBlue* citation handbook asserting potential copyright and trademark violations).

[21] BLUEBOOK, *supra* note 20.

series of West's *Federal Reporter*, which comprises federal appellate opinions from 1925 to 1993, is "F.2d."[22])

The result of the *Bluebook*-dominated legal-citation ecosystem is that volume-reporter-page number citations assume a relatively standardized form, even if surrounding information like the case caption may be presented inconsistently. This, in turn, means that citations to caselaw—or phrases that look like citations to caselaw—are relatively easy to extract programmatically from documents in an automated fashion. (At least, case citations are easier to extract programmatically than the broader and more indeterminate universe of citations to any conceivable published source.)

Next, Subsection A reports how we extracted citations from our dataset and attempted to determine whether they pointed to genuine caselaw.[23] Subsection B explains why that determination—whether a citation points to genuine caselaw—is one of devilish and underappreciated complexity. Subsection C synthesizes the preceding sections to explain why the concept of the "hallucination" is underspecified and probably impossible to ascertain from inspecting suspect text alone.

A. Practical Challenges (and Technical Details)

1. The Data

We assembled an original dataset of 2,669,573 legal filings spanning 399,283 cases from four jurisdictions between 2015 and 2025. We collected these data from one state court, Oklahoma County District Court, and three federal district courts: the Northern District of Oklahoma, the Western District of Oklahoma, and the Southern District of Texas.

Although federal court filings are accessible to the public through the Public Access to Court Electronic Records (PACER) service, PACER charges a fee of 10¢ per page.[24] We obtained data from these federal jurisdictions pursuant to a PACER fee waiver for academic research, the terms of which limit our ability to share or republish the documents obtained. PACER fees have been the subject of much vexation in civil

---

[22] *E.g.*, 1 THE FEDERAL REPORTER: SECOND SERIES (1925). *See also* BLUEBOOK, *supra* note 20, at T1.1 (specifying the appropriate abbreviations for federal reporters).

[23] This section reproduces information that we present in considerably more detail in an accompanying technical paper written for a computer-science audience. We are currently preparing that paper for publication in a computer-science venue. Liu et al., *supra* note 9.

[24] *PACER Pricing: How Fees Work*, PUB. ACCESS TO CT. ELEC. RECS. https://pacer.uscourts.gov/pacer-pricing-how-fees-work (last visited July 13, 2026).

society[25]; the proposed federal Open Courts Act of 2026 would establish a new, centralized system for which "[n]o fees may be charged."[26] (We discuss PACER-related discontents and similar access-to-justice problems at length in Part III.A.) We obtained explicit permission to scrape dockets from the Oklahoma State Courts Network.

From the Western and Northern Districts of Oklahoma, we scraped every case filed from 2015 to 2025, obtaining 22,312 and 10,996 cases, respectively. For the Southern District of Texas, we scraped a random sample of 2,000 cases for each year from 2025 to 2021, a sample of 9,990 cases from 2022, and every case from 2023 to 2025, for a total of 62,030 cases. For Oklahoma County, we scraped every case filed in the first nine days of each month (i.e., on January 1–9, February 1–9, etc.) from 2015 to 2025, retrieving 303,945 cases. We saved all accessible docket entries,[27] amassing a corpus of 2,669,573 documents across 399,283 cases. Scraping was conducted in July and August 2025.

Table 1: The Number of Cases, Documents, and Citations Processed by Jurisdiction

Jurisdiction	# Cases	# Documents	# Citations
Oklahoma County	303,945	809,597	448,331
S.D. Tex.	62,030	963,511	1,000,169
W.D. Okla.	22,312	603,784	1,113,351
N.D. Okla.	10,996	292,681	561,119
Total	399,283	2,669,573	3,122,970

[25] *See* Letter from Fix the Court et al. to Sens. Durbin and Grassley (June 15, 2026), https://fixthecourt.com/wp-content/uploads/2026/06/Group-letter-on-the-Open-Courts-Act-6.15.26.pdf.

[26] Open Courts Act of 2026, S. 4667, 119th Cong. §§ 2, 3 (2026).

[27] Some documents were sealed or otherwise inaccessible to the public.

2. The Tool

This section offers a brief overview of the original "Hallucite" tool, which will soon be made publicly available and full technical details of which are described at length in the accompanying technical paper.[28]

a) Document Processing and Citation Extraction

The first step in our analysis involved processing scraped documents and extracting probable citations from them. Readers interested in the particulars of this process should consult the technical paper, which describes it at length.

For our present purposes, the main takeaway is that several phenomena impede automated textual analysis of legal filings. Those issues include the fact that the documents do not always contain embedded text[29] and may therefore require fallible optical character recognition (OCR) analysis[30] to become machine-readable. Furthermore, formatting conventions in legal documents, such as line numbers, footnotes, headers, and tables of authorities, hamper the extraction of citations; hence, extensive data cleaning and preprocessing is essential for effective citation analysis. Once we cleaned our data, we extracted legal citations using a computer-science technique called regular-expression matching,[31] building on the Free Law Project's Eyecite Citation Extractor.[32]

---

[28] Liu et al., *supra* note 9.

[29] Some documents are "native" or "born digital" documents that were created electronically and have selectable, embedded text. Other documents are printed and scanned. All of the documents from Oklahoma County were scanned. Roughly a third of documents from the three federal district courts were scanned, although that proportion declined over time.

[30] *See* Daniel Lopresti, *Optical Character Recognition Errors and Their Effects on Natural Language Processing*, *in* PROCEEDINGS OF THE SECOND WORKSHOP ON ANALYTICS FOR NOISY UNSTRUCTURED TEXT DATA 9, 9 (2008) ("Errors are unavoidable in advanced computer vision applications such as optical character recognition, and the noise induced by these errors presents a serious challenge to downstream processes that attempt to make use of such data."); Thomas Hegghammer, *OCR with Tesseract, Amazon Textract, and Google Document AI: A Benchmarking Experiment*, 5 J. COMPUT. SOC. SCI. 861, 862 (2022) (comparing three OCR technologies and finding that "OCR is a technology still in the making, and available software provides varying levels of accuracy.").

[31] Ken Thompson, *Regular Expression Search Algorithm*, 11 COMMC'NS ACM 419, 419 (1968) (describing a "method for locating specific character strings embedded in character text.").

[32] *Eyecite Citation Extractor*, FREE L. PROJECT, https://free.law/projects/eyecite/ (last visited July 13, 2026).

b) Building a Citation Database

After extracting everything in our dataset that resembled a citation, we had to check it against a database of authoritative citations. Once again, this is a task easier said than done. To our knowledge, no complete and authoritative database exists that is publicly accessible.[33] As a result, we compiled our own database of canonical case citations, which we use to match references to a volume, reporter, and page number with a corresponding case.

(1) What's in the Database?

Our database is built upon data from the Free Law Project's CourtListener intiative, which has compiled over 15 million citations.[34] The CourtListener dataset is an invaluable undertaking, and we are indebted to its compilers; however, gaps in data coverage remain. For instance, until early 2026, its coverage of the U.S. Reports concluded around 2015, so some landmark contemporary decisions such as *Dobbs v. Jackson Women's Health Organization* were missing from the base dataset.[35] At the time of writing, significant federal district court cases are still missing, such as *NetChoice v. Moody*, 546 F. Supp. 3d 1082 (N.D. Fla. 2021).[36] We supplemented the CourtListener data with information from the commercial database Leagle,[37] the Supreme Court's official website,[38] the Supreme Court Database,[39] and, for Oklahoma-specific reporters, the Oklahoma State Courts Network (OSCN) website.[40] Without these data supplements, 2.4% of citations would have been incorrectly flagged as invalid.

---

[33] *See generally* Susan Nevelow Mart, *The Algorithm as a Human Artifact: Implications for Legal [Re]Search*, 109 LAW LIBR. J. 387 (2017).

[34] *Data Coverage — What Case Law Does CourtListener Have?* , COURTLISTENER, https://www.courtlistener.com/help/coverage/opinions/ (last visited July 27, 2026) ("Our database encompasses more than 99% of all precedential legal case law published in the United States.").

[35] 597 U.S. 215 (2022). We speculate that until recently, CourtListener did not have U.S. citations outside of the officially published volumes, which, as of July 2026, go only up to Volume 587.

[36] *Unable to Find Citation "546 F. Supp. 3d 1082"*, COURTLISTENER, https://www.courtlistener.com/c/f-supp-3d/546/1082/ (last accessed July 29, 2026).

[37] LEAGLE, https://www.leagle.com/ (last accessed July 27, 2026).

[38] *Opinions*, SUP. CT. U.S., https://www.supremecourt.gov/opinions/opinions.aspx (last accessed July 27, 2026).

[39] *The Supreme Court Database*, WASH. U.L., http://scdb.wustl.edu/ (last accessed July 27, 2026).

[40] Okla. State Cts. Network, https://www.oscn.net/home/ (last accessed July 27, 2026).

(2) What's *Not* in the Database?

Proprietary caselaw citations—chiefly to documents in the Westlaw and Lexis databases—represent a large gap in our database. Regrettably, these proprietary citations also represent a significant portion of the citations we detected: 11.3% and 2.1% of the citations we found pointed to Westlaw's and Lexis's databases, respectively. These sites' terms of service forbid any form of automated access,[41] and the vast majority of proprietary citations cannot be found in open-source datasets. As a result, we could not validate citations to Westlaw and Lexis in an automated way. Our validation analysis omits them. Other significant gaps include tax court cases, bankruptcy cases, and cases not selected for official publication.[42]

c) Citation Validation

When a citation is identified in a document, validation proceeds in three steps. First, the system queries the citation database for an entry with matching volume, reporter, and page number. If no such entry exists, then the citation does not resolve to an existing case and is considered invalid. If a case is found, then the case name in the database must be compared to the case name in the document to ensure that they match. For instance, consider a hypothetical citation "Roe v. Wade, 5 U.S. 137." The reporter citation "5 U.S. 137" is a valid reference to *Marbury v. Madison*, but the entire citation is invalid because the given case name, "Roe v. Wade," does not match the actual case name. AI systems appear particularly prone to generate valid reporter citations along with an unrelated case name: In a sample of 186 invalid citations known to be AI generated, we found 27% of them resolved to a real case but had the wrong case name.

The second step is to extract the case name from the document text. This involves stepping backward from the reporter citation, one word at a time, until the beginning of the case name has been reached. A typical

[41] *General Terms and Conditions*, LEXISNEXIS, https://www.lexisnexis.com/en-us/terms/general/default.page (last visited July 13, 2026) ("You may not . . . [a]ccess or use the Online Services via mechanical, programmatic, robotic, scripted or any other automated means"); *General Terms and Conditions (Version 5.1)*, 3 (2025), https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/other/general-terms-us-q225.pdf ("[Y]ou must not . . . mine, scrape, index, or otherwise automatically access, collect, copy, download or record the Property . . . .").

[42] "Unpublished" cases appear both in the *Federal Appendix* and in Federal Reporter volumes but are not available in publicly accessible databases including CourtListener, the Caselaw Access Project, or Google Scholar. It is nonetheless permitted to cite such cases. FED. R. APP. P. 32.1.

citation occurs after a complete sentence, but citations are also used within sentences, in footnotes, within citation parentheticals, in citation strings, and even within headings. Hallucite applies a variety of heuristics to try to pull out the full case name without dredging up any extraneous surrounding text.

Third, the case name extracted from the text and the case name retrieved from the database are compared to ensure that they match. The challenge is that there is no single canonical name for a case. The landmark case of *Brown v. Board of Education* appears nineteen times in our dataset, spelled six different ways.[43] Known abbreviations, acronyms, and initialisms must be replaced by their fully expanded form. Common suffixes like "LLC," "Co.," and "Inc." are sometimes omitted, or surnames are used in lieu of full names. Hallucite flexibly handles the wide variety of case names that occur in real-world citing practices, including minor typos that afflicted 4% of citations in our dataset. Hallucite also handles minor typos present in 2% of reporter citations in our dataset, such as a page number that is missing a digit.

Hallucite's validation process ultimately categorizes citations as valid or invalid. As explained above, a citation that Hallucite classifies as "valid" is not necessarily formally perfect. Because Hallucite flexibly handles minor typographical variations in case names, for example, a citation that it classifies as valid may nevertheless contain errors within its tolerance. When Hallucite classifies a citation as "invalid," by contrast, this means that the citation could not be resolved to an existing case even with tolerance for some orthographical errors.

A significant number of citations could not be judged as either valid or invalid and had to be excluded from analysis. Nearly 18% of the citations we extracted fall into this category. The vast majority of these were to proprietary reporters owned by West and Lexis. Other citations were missing a case caption, a party name, or a portion of the reporter citation. All of the results reported in the remainder of this Article exclude these citations.

d) Benchmarking and Performance

We evaluated Hallucite's performance on a set of 566 citations extracted from 40 documents taken from real-world cases identified in

[43] 347 U.S. 483 (1954). Variations included Brown v. Bd. of Educ. (8) , Brown v. Board of Education (7), Brown v. Bd. of Ed. of Topeka (1), Brown v. Board of Ed. (1), Brown v. Bd. of Ed. of Topeka, Shawnee Cnty., Kan. (1), and Brown v. Board of Education of Topeka (1).

Damien Charlotin's tracker of cases involving AI-generated legal citations.[44] We manually verified the existence of each citation using Westlaw and confirmed that 381 (67%) of the citations are real while 185 (33%) did not resolve to any existing case. Unfortunately, 101 of these citations—78 of the 185 invalid citations and 23 of the 381 real ones—point to proprietary reporters that Hallucite cannot verify. Restricting to the 465 remaining citations, the system achieves an accuracy rate of 99.8%, correctly classifying 464 of 465. The one misclassified citation is of a real case that is not officially published in the Federal Reporter and hence does not appear in open-source databases.[45] These "unpublished" cases may be cited in briefs even though they carry no precedential value.[46]

3. The Results

Across two and a half million citations extracted from two and a half million filings from four jurisdictions over 2015–2025, we find no meaningful increase in the proportion of invalid citations after 2022, when generative AI became prevalent. Adjusting for jurisdiction and OCR status,[47] 0.92% of citations filed before December 1, 2022[48] were categorized as invalid, compared to 0.90% of citations filed after. Through careful manual review of a sample of these citations, we find that the vast majority do refer to real citations but couldn't be automatically validated. A small

[44] Charlotin, *supra* note 9.

[45] The unpublished citation is Tucker v. Tucker, 203 F.3d 832 (9th Cir. 1999), cited in Objection to the Magistrate Judge's Report and Recommendations at 20, Arnaoudoff v. Tivity Health Inc., No. CV-23-01510-PHX-DJH, 2025 WL 761452 (D. Ariz. Mar. 11, 2025), Dkt. No. 63. This citation to *Tucker* is not available in CourtListener. *Unable to Find Citation "203 F.3d 832"*, COURTLISTENER, https://www.courtlistener.com/c/f3d/203/832/ (last accessed July 30, 2026). Caselaw Access Project has made available a scan of the entire volume, but the 832nd page is entirely redacted, along with all other pages listing unpublished cases. *203 F.3d*, CASELAW ACCESS PROJECT, https://static.case.law/f3d/203.pdf (last accessed July 30, 2026).

[46] *See* FED. R. APP. P. 32.1 (allowing citation of cases "designated as 'unpublished,' 'not for publication,' 'non-precedential,' 'not precedent,' or the like").

[47] Documents from Southern District of Texas had lower invalid-citation rates, and their proportion in our dataset increased over time. OCR documents have higher invalid-citation rates, and their proportion declined over time. We weighted the data by jurisdiction and OCR status to correct for this shift in composition.

[48] On November 30, 2022, ChatGPT was the first generative AI chatbot to be publicly released. *Introducing ChatGPT*, OPENAI, https://openai.com/index/chatgpt/ (last accessed July 30, 2026).

minority, however, are genuinely invalid, and may include hallucinated citations that have gone unnoticed.

a) Real Citations That Failed To Validate

We selected a random sample of citations deemed invalid by Hallucite and attempted to manually resolve the citation through searching on Westlaw and looking at the broader document text. Over 90% of these citations turn out to resolve to real cases but could not be automatically validated for a variety for a variety of reasons: OCR artifacts, formatting errors,[49] abbreviations unknown to Hallucite,[50] cases with multiple case captions,[51] unpublished cases,[52] and scrivener's errors.[53]

These results show that there are serious hurdles to improving automated validation of legal citations. Over time, data cleaning and Hallucite's knowledge of abbreviations could improve, and with great effort, a more comprehensive open-source legal database could be compiled. But nearly half of Hallucite-invalid citations appear to be due to human oversight in formatting and spelling. An automated system could be more lenient with such defects, but that inevitably increases the risk that a truly invalid citation is wrongly classified as valid.

b) Truly Invalid Citations

The vast majority of citations that were deemed invalid by Hallucite referred to real cases even if they were formally defective in some fashion. A small remainder were truly invalid and did not refer to any case we could find on Westlaw. It is difficult to know from the text alone whether these

[49] E.g., short form citations, such as "*Iqbal*, 556 U.S. at 678," sometimes omitted the "at," becoming an invalid full case citation.

[50] E.g., Consolidated Rail Corp. v. Gottshall, 512 U.S. 532 (1994), was cited as "Conrail v. Gottshall" and Hallucite could not match the two names. Other unknown abbreviations include "UPS" for the "United Parcel Service," "AMTRAK" for the "National Railroad Passenger Corporation," and "USSA" for the "United Services Automobile Association."

[51] E.g., Haga v. Astrue, 482 F.3d 1205 (10th Cir. 2007), was renamed from Haga v. Barnhart due the transition from Jo Anne B. Barnhart to Michael J. Astrue as Commissioner of the Social Security Administration. CourtListener only has the case with the older name. *Found More than One Result for "482 F.3d 1205"*, COURTLISTENER, https://www.courtlistener.com/c/f3d/482/1205/ (last accessed July 30, 2026).

[52] *See, e.g.*, *supra* note 47.

[53] Some were simple misspellings: Stump v. Sparkman, 435 U.S. 349 (1978), was cited as "Stump v. Steward" and United States v. McVeigh, 153 F.3d 1166 (10th Cir. 1998) was cited as "United States v. McVay." Some reporter citation typos could be explained by the misplacement of a finger, such as 540 F.3d 333 being improperly cited as "540 F.3d 222." More complex examples are discussed in the technical manuscript. Liu et al., *supra* note 9.

invalid citations involved AI misuse, which we discuss later below in Part C.2.b). But some factors are suggestive of an AI-generated citation.

Consider this excerpt from a complaint in our dataset: "Courts have extended Monell liability to public universities when student organizations engage in discriminatory or unconstitutional conduct with university knowledge. See Foss v. Arizona Board of Regents, 82 F.4th 1372 (9th Cir. 2023)." The only case with this name is a 2019 case in the Arizona Court of Appeals.[54] 82 F.4th 1372 refers to the second page of a 2023 opinion from the Federal Circuit.[55] This citation is invalid.

While we are not certain, this citation may have been AI-generated. First, the case caption appears superficially relevant: the preceding proposition and the complaint concern public universities, and the cited case has Arizona Board of Regents as a party. In our anecdotal experience, hallucinated citations tend to include case captions that appear relevant. Second, even if the citation was intended to refer to the case in the Arizona Court of Appeals, that case does not support the proposition it is being cited for. The case has nothing to do with *Monell* liability; it is about who may be entitled to in-state tuition benefits.[56] Third, the complaint was filed in 2025 by a *pro se* litigant, a graduate student in a school of social work who likely lacks formal legal training.

None of these are dispositive, but together we believe they are suggestive of AI use. Alternatively, they may be the result of sloppy drafting. As the judge in the case put it, the "pleadings have not been a model of clarity."[57] This example highlights the difficulty with determining the provenance of a citation from text alone, even with careful human review. In fact, we conjecture in Part C.2.b) that it is impossible.

c) Valid Citations May Be Formally Defective

We found that 4% of valid citations had a minor typo in a party name, and a further 2% of citations had a minor typo in the volume number, reporter name, or page number. That means roughly one in every sixteen of the citations that we classify as "valid" does not conform exactly to a

[54] Foss v. Arizona Bd. of Regents, No. 1 CA-CV 18-0781, 2019 WL 5801690 (Ariz. Ct. App. Nov. 7, 2019).

[55] Schwendimann v. Neenah, Inc., 82 F.4th 1371 (Fed. Cir. 2023).

[56] Foss, 2019 WL 5801690, at *1 ("At issue here is whether three non-resident students are entitled to recoup the cash difference between in-state and out-of-state tuition for the 2017-2018 academic year because a discrete group of undocumented aliens received in-state tuition rates to attend Arizona's public universities during that period.").

[57] Chowdhury v. Univ. of Houston, No. 4:25-CV-467, 2026 WL 759774, at *1 (S.D. Tex. Mar. 18, 2026). There is no discussion in the opinion about possible AI use. *Id.*

standard-form citation that denotes the cited case. (In our terminology, such citations are "referentially valid," but not "*strictly* referentially valid.")[58] These results do not demonstrate a pandemic of AI-assisted legal malpractice. Rather, they demonstrate the theoretical challenges of specifying precisely when a defective citation is cause for greater concern.

B. Theoretical Challenges

Our technical challenges resulted in part from larger theoretical challenges about what it means for text to be a citation, and what it means for a citation to be valid or invalid. The following subsections specify that a citation may be invalid because it is "referentially invalid" (it fails to denote an existing legal authority) and/or because it is "propositionally invalid" (it makes a legally indefensible characterization of an authority). A "hallucinated citation," for our purposes, is an ostensible citation that lacks *either* referential or propositional validity and was introduced by the use of generative artificial intelligence.

1. Defining "Citation"

Calling text an "invalid citation" presupposes that it is a "citation" in the first place. Plenty of phrases—"To Whom It May Concern", "The parties shall meet and confer on June 5", etc.—fail to function as legal citations not because they are citations that fail, but because they are not citations at all. Beneath this obvious observation is a complicated question: what makes particular text a citation? What does it mean to say that a hallucination "is" a "citation" yet not *actually* a citation?

These are difficult ontological issues, and we approached them pragmatically. We decided to treat text as a "citation" when it bore a formal resemblance to a standard-form Bluebook case citation.[59] This format is the customary form of case citations in legal practice in the United States; indeed, courts may even specifically instruct parties to adhere to Bluebook format.[60] The Bluebook case-citation format also has the advantage of being standardized enough to detect reliably through regular-expression matching. Of course, permissible citations in legal materials may not always look like a standard case citation. For instance, "*Id.*" can be a valid case citation, but it operates deictically instead of as a durable pointer to a

[58] *See infra*, Part I.B.2.a).

[59] Liu et al., *supra* note 9.

[60] *See, e.g.*, *Citation Formats*, U.S. CT. FED. CLAIMS (2019), https://www.uscfc.uscourts.gov/sites/cfc/files/citation_formats.pdf.

particular source; its referent is determined by the surrounding context. We could have chosen to make classification as a "citation" depend on a speaker's intent or a listener's understanding—and there are surely strong philosophical arguments for doing so—but examining the formal characteristics of the text was the approach we deemed administrable at scale.

2. Defining Citation "Validity"

The next step is to specify what makes a citation "valid." In the sense we use the term, a "valid citation" is a reference to a legal decision made in a manner consistent with an attorney's standard of care. That standard of care entails, at minimum, not citing "phony case law"[61] and not misrepresenting the proposition for which a case stands.[62]

A citation's validity encompasses at least two distinct properties.[63] First, it must be "referentially valid": it must actually refer to an existing legal authority. Second, it must be "propositionally valid": it must state a proposition that, if not unequivocally true, is at least "legally tenable."[64] The Ninth Circuit recently articulated a similar distinction between "fabrications"—"instances in which the generative AI tool provides cases or quotations that do not exist at all" and "inaccuracies"—instances in which "[t]he generative AI tool might cite to real authorities but provide an answer that is legally or factually inaccurate or not supported by the citation."[65] Prior studies of AI hallucinations have focused on propositional validity rather than referential validity (see *infra* Subsection (a)(1)). Our study, by contrast, focused on referential validity, which proved exceedingly difficult to formalize.

---

[61] Turnage v. Associated Bank, N.A., No. 25-CV-3004 (ECT/DTS), 2025 WL 3052638, at *3 (D. Minn. Sept. 12, 2025), aff'd, No. 25-3092, 2025 WL 4694563 (8th Cir. Nov. 25, 2025) ("Every filing in a federal court that contains citations to phony case law amounts to a violation of Rule 11(b) of the Federal Rules of Civil Procedure."); ABA Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024) ("[O]utput from a [generative]AI tool must be carefully reviewed to ensure that the assertions made to the court are not false.").

[62] *See* Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 460 (S.D.N.Y. 2023) (citing United States v. Fernandez, 516 Fed. App'x 34, 36 & n.2 (2d Cir. 2013)).

[63] Cf. Cheng-chi (Kirin) Chang, *Hallucinated Authority: AI Citations as Reckless Misrepresentation* , 115 GEO. L.J. ONLINE (forthcoming 2026) at 16 (Mar. 24, 2026) (unpublished manuscript), https://papers.ssrn.com/abstract=6464098 (discussing verification as requiring "both that [a] cited case exists and that it stands for the proposition cited").

[64] *Cf.* Cooter & Gell v. Hartmarx Corp., 496 U.S. 384, 393 (1990).

[65] Malkeet Lnu v. Blanche, 177 F.4th 1014, 1024–25 (9th Cir. 2026).

a) Referential Validity

Referential validity means that the citation must refer to an actual case. Although that definition sounds straightforward, "referring to" is a complicated linguistic phenomenon that scholars across disciplines have found challenging to formalize.[66] We use reference to describe the meaning a reader apprehends in context, rather than the meaning a speaker may have intended (which may or may not be the same).[67] For our purposes, a citation "refers to" a case—and thus is referentially valid—when a reasonable reader would ascertain a unique, existing case that corresponds to the citation. In other words, we do not evaluate whether the *speaker* is referring, but instead whether *the citation* refers.[68]

Strict referential validity requires that the reference conform exactly to a standard-form citation that denotes the cited case. "State v. Shack, 277 A.2d 369 (N.J. 1971)" satisfies strict referential validity, as does "State v. Shack, 58 N.J. 297 (1971)." However, citations may be referentially valid without satisfying strict referential validity. "State v. Shacke, 277 A.2d 368 (N.J. 1971)" makes a minor misspelling of a party's name and lists a starting page that is one page too early. This citation is not an exact match for the standard-form citation for *State v. Shack*. But absent unusual circumstances,[69] the slightly malformed citation will unequivocally refer to the case; almost any lawyer who encounters this citation in context will, with little effort, be able to find the case to which it refers and infer the reasons why the citation fell short of perfection. Courts routinely process citations that, while formally defective, still communicate a reference.[70] Of course, in marginal cases, referential validity may be hard to assess. What sorts of readers

---

[66] *See generally* Eliot Michaelson, *Reference*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY, https://plato.stanford.edu/entries/reference/ (Feb. 12, 2024).

[67] *Cf. generally* H. P. Grice, *Utterer's Meaning, Sentence-Meaning, and Word-Meaning*, 2 *in* PHILOSOPHY, LANGUAGE, AND ARTIFICIAL INTELLIGENCE 49 (JACK KULAS, JAMES H. FETZER, & TERRY L. RANKIN EDS., 1968), http://link.springer.com/10.1007/978-94-009-2727-8_2.

[68] Matthew Mandelkern & Tal Linzen, *Do Language Models' Words Refer?*, 50 COMPUT. LINGUIST. 1191, 1192 (2024).

[69] E.g., a situation in which a *different* case named *State v. Shacke* was itself published on a nearby page of the same reporter volume.

[70] *See* L. Off. of Shawn Shearer, P.C. v. Iowa Dist. Ct. for Fremont Cnty., 27 N.W.3d 806, 815 n.3 (Iowa 2025) (inferring "most likely" intent behind typographical error in citation to statute); Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 n.4 (Ind. Ct. App. 2026) ("We are mindful that the increasingly prevalent use of generative artificial intelligence (AI) has led many courts to sanction attorneys whose filings included fake case citations. But we also keep in mind that this Court has typically shown grace to practitioners whose citations—though incorrect—appeared to be the result of inadvertent typos."); *infra* Part II.C.

constitute the relevant audience, and what percentage them must be able to ascertain the cited case to preserve referential validity? 50%? 100%? 75%? We doubt that this question has a formal answer; we certainly do not provide the formula.

The stakes of referential validity depend on historical and professional context. Under the strict pleading standards formerly in place in English courts, for example, a lawyer might risk forfeiting a claim if his pleading or replication contained any sort of formal defect.[71] In the present-day United States, a lawyer would almost certainly not face professional sanction for submitting a citation that—due to a typographical error, for example—failed strict referential validity but still satisfied referential validity in the broader sense. Indeed, courts have specific procedures for correcting scrivener's errors in legal documents.[72]

(1) Existing Studies Tend to Elide Referential Validity

Prior analyses of hallucinated legal citations have largely avoided the problem of referential validity through experimental design. A pioneering study by Dahl et al. uses a methodology that bypasses difficult referential problems.[73] Dahl et al. measure LLMs' performance on a series of legal tasks of varying complexity, some of which deal explicitly with citation and existence, such as "Is {case} a real case?" and "What is the citation for {case}?"[74] However, all of the tasks in Dahl et al.'s paper *supply* a case that the LLMs' response is to be "about."[75] This experimental design fixes the case that serves a reference point and measures whether the system produces accurate factual information about that case, such as its citation, its authoring judge, or its holding.[76] This format sidesteps the difficult antecedent question of referential validity because it removes any occasion to ask, "what case, if any, does this AI output refer to?" Instead, Dahl et al.

---

[71] William Feilden Craies, *Pleading*, *in* 21 ENCYCLOPÆDIA BRITANNICA (1911) https://en.wikisource.org/wiki/1911_Encyclop%C3%A6dia_Britannica/Pleading (last visited June 2, 2026) ("As an extreme instance, a learned judge in the 19th century challenged a pleading for putting the year without adding A.D., on the ground that 'non constat that A.C. might not be intended.'"). *Cf. also, e.g.*, Edward Crogate's Case, 77 E.R. 574; (1608) 8 Co. Rep. 66; [1608] 1 WLUK 167, which was satirized in GEORGE HAYES & EDMUND MACRORY, HAYESIANA 7 [Jeux d'esprit by G. Hayes, Ed. with a Biogr. Preface by E. Macrory]. (1892), https://catalog.hathitrust.org/Record/103401029.

[72] United States v. Cintron-Echautegui, 604 F.3d 1, 4 n.2 (1st Cir. 2010).

[73] Matthew Dahl et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, 16 J. LEG. ANAL. 64 (2024).

[74] *Id.* at 68–70.

[75] *Id.* at 69 ("We use only real cases in our prompts . . . .").

[76] *Id.* at 68.

stipulate the case that is to be treated as the referent and measure whether the output is propositionally valid with respect to that referent.[77]

Another study by Magesh et al. analyzes the output of AI legal assistants in terms of its "correctness" and "groundedness." Correctness, the authors explain, represents a straightforward inquiry into propositional validity: is the AI-generated assertion "factually correct and relevant" to the query?[78] As for "groundedness," the authors explain, "[a] response is *grounded* if the key factual propositions in its response make valid references to relevant legal documents."[79] Our concept of referential validity differs from Magesh et al.'s groundedness:[80] Magesh et al. do not specify what it means for an AI assistant to "make [a] valid reference to [a] legal document[]" in the first place. In the context of their study, this lacuna is perfectly reasonable: Magesh et al. devote their analysis to retrieval-augmented generation (RAG) tools, which respond to queries by consulting an external database of information.[81] While a LLM without RAG would be limited to the information in its training data, a RAG-based system can consult "a legal database, retrieve the relevant metadata, and then provide that to the LLM, which would use the result to provide the user a response to their query."[82] RAG legal assistants furnish documents, with links to specific materials in legal databases, in their responses to users' queries.

Thus, by limiting their inquiry to RAG-based tools, Magesh et al. can take advantage of a readymade scaffolding that establishes (a) whether and when an output is referring to a legal authority, and (b) what legal authority is being referred to. Magesh et al. evaluate a response's groundedness with respect to the documents "provide[d]" and "cite[d]" by the tool—these are the documents being referred to.[83] Because RAG-based tools connect to an authoritative database — say, of legal materials — they can at least help ensure that *linked* citations correspond only to real cases that appear in the authoritative database. Magesh et al. note that Lexis's AI tool may in this way avoid linking to nonexistent legal documents, and could in that narrow

[77] In an appendix, Dahl et al. describe the results of "a supplemental analysis where [the authors] repeat the existence query, but using fake cases instead of real cases." *Id.* at 61.
[78] Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. EMPIR. LEG. STUD. 216, 220–21 (2025).
[79] *Id.* at 221.
[80] *See infra* Subsection (b) for more discussion of the differences.
[81] *Id.* at 222.
[82] *Id.* at 219.
[83] *Id.* at 221.

sense be "hallucination-free" (but other commentators have suggested this is not in fact the case).[84]

Recent work by Patty Liu et al. confronts referential validity more directly. The authors prompt several LLMs to generate legal documents, then attempt to verify the case citations in the generated documents using the CourtListener and Westlaw databases.[85] The authors explain that they "classify a citation as hallucinated if it does not correspond to a real case or if the generated case name does not match the name associated with the reporter reference in Westlaw."[86] This definition of hallucination appears to impose, in our terminology, a requirement of strict referential validity.[87] The authors assert that their "definition captures only the most obvious hallucinations and provides a conservative lower bound on overall prevalence of such hallucinations."[88] In fact, however, our results suggest the opposite: their definition would surely overestimate invalid citations if applied to a corpus of human-authored legal documents, because many human citations are referentially valid despite not being strictly referentially valid.[89]

As an example of a "hallucination," Patty Liu et al. provide the text "*In re Grand Jury Subpoenas*, 995 F.2d 348, 352–53 (2d Cir. 1993)." They explain, "the citation *995 F.2d 348* . . . exists, but refers to the case *Malcolm v. National Gypsum Co.* We classify it as hallucinated due to the inconsistency."[90] This "hallucination" fails strict referential validity: there is no case with that exact citation. According to a Westlaw search, there isn't even a Second Circuit case from 1993 called *In re Grand Jury Subpoenas*, although there is one from 2003.[91] But *Malcolm v. National Gypsum Co.* is a real Second Circuit case from 1993, and it begins on page 346 (not page 348)

[84] *Id.* at 221; But see *Id.* at 234 n.8 (Noting that "there is some evidence that Lexis+ AI does not succeed even by this metric. McGreel [2024] reports instances of Lexis+ AI citing cases decided in 2025.").

[85] Liu, Stammbach, and Henderson, *supra* note 3 at 3.

[86] *Id.*

[87] We say "appears to" because elsewhere in the article, Patty Liu et al. use phrasing that suggests that slight errors may not warrant classifying a citation as a hallucination. *See id.* at 5 ("We generate non-existent citations by altering the reporter volume, reporter abbreviation, or page number to *implausible* values." (emphasis added)).

[88] Liu, Stammbach, and Henderson, *supra* note 3 at 4.

[89] Recall our finding that "at least one in every twenty-five valid citations fails to satisfy" strict referential validity. *Supra* Part I.A.3.

[90] *Id.* at 3–4.

[91] In re Grand Jury Subpoenas Dated Mar. 19, 2002 & Aug. 2, 2002, 318 F.3d 379 (2d Cir. 2003).

of the 995th volume of the Second Series of the Federal Reporter, and the pincited pages 352 and 353 are pages of that decision.[92]

While it is entirely reasonable for Patty Liu et al. to classify this example as a hallucination, the reasonableness of that decision depends on their knowledge that it was produced by AI. There is probably a context in which "*In re Grand Jury Subpoenas*, 995 F.2d 348, 352–53 (2d Cir. 1993)" would be referentially valid. For example, if a reader encountered this exact text in a filing not known to be AI-authored, and the text appeared alongside discussion of the legal standard for consolidating asbestos exposure cases (which is what pages 352 and 353 of *Malcolm* discuss), then the text probably *would* successfully refer to a particular portion of that particular case.[93] Thus, while Patty Liu et al.'s definition of "hallucination" makes sense to apply to their dataset of filings known to be generated by LLMs, it would be untenable to apply to a sample of citations from actual legal filings whose provenance is necessarily unknown. Classifying as a hallucination every citation that fails strict referential validity—or even every citation that is as far from formal perfection as the *In re Grand Jury Subpoenas* example—would ensnare plenty of human errors that might still, in context, successfully communicate a reference.[94]

(2) Referential Validity Remains an Unavoidable Problem

As the preceding literature review demonstrates, assessing a citation's referential validity is, in some contexts, relatively straightforward. But referential validity will remain a major practical problem in legal practice for least three reasons.

First, even supposing that RAG-based systems "solve" referential validity by always linking to real documents in an authoritative database—such that it is always clear whether these systems are referring to a case and what case they are referring to—not all legal filings will be drafted with these tools. Plenty of litigants file documents generated by LLMs that output citations, or apparent citations, without linking to any supporting documents.

Second, a legal filing does not have the same user interface as a chatbot. It may be possible for a chatbot to provide inline links to any

[92] Malcolm v. Nat'l Gypsum Co., 995 F.2d 346 (2d Cir. 1993).

[93] Experienced readers can sometimes tell when a lawyer appears to have "combined . . . two case citations without noticing his mistake," Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 (Ind. Ct. App. 2026), yielding a citation that is referentially valid but fails *strict* validity.

[94] Liu et al., *supra* note 9.

documents it may reference. However, a plain-text citation—especially one in a hardcopy brief—cannot. Instead, judges and law clerks must plug citations into a legal database to find the cited sources, and if a filing contains an apparent citation that fails strict referential validity, it will require some guesswork to infer its referent.

Third, links are themselves fallible. Among other things, a link to a purported source might lead to an "Error: File Not Found" page; a functioning hyperlink might contain an HREF link to a resource that does not correspond to the source linked in the hyperlink text field; or a particular context or interface might make it difficult to decipher which portions of an utterance refer to a linked document, and which portions do not.[95]

Because our study examined plaintext citations from legal filings, we could not avoid the problem of referential validity in the ways that Dahl et al., Magesh et al., and Patty Liu et al. could. While Dahl et al. were *themselves* referring to particular cases in executing their experiment, our analysis involved *litigants'* references to cases; this meant that we could not stipulate a referent and instead had to determine what cases the litigants were referring to, and when. Unlike Magesh et al., we were not interfacing with a system that provided links and supporting documentation that we could treat as the grounding for a given assertion. Instead, we had to determine whether the citations themselves effectively pointed to that supporting documentation. Finally, unlike Patty Liu et al., we were not analyzing citations we knew to be AI-generated, which meant that we could not define "hallucination" to include any apparent citation that lacked strict referential validity, because doing so would have ensnared countless human errors that made no material impact on a citation's efficacy. As a result, our research is the first in this line to take on the difficult and essential problem of specifying when text refers to anything at all.

b) Propositional Validity

Propositional validity means that a citation must not misrepresent the proposition for which a cited authority stands. Our standard for propositional validity focuses on a reader's assessment rather than a speaker's intent. What matters is not whether a writer intended to misrepresent a cited authority, but rather whether an objective or

[95] As an example of the lattermost possibility, consider a sentence that asserts several propositions and is followed by two hyperlinks, presented as, "(Source 1: X)" and "(Source 2: Y)." This presentation of information may not unambiguously convey which source, if any, supports which assertion.

reasonable reader — in our context, of a case citation in a legal filing — would conclude that the authority is misrepresented.

Misrepresentation may be *explicit*: a brief might cite a case for a proposition incompatible with its holding. Misrepresentation may also be *implicit*. For example, a brief filed in 2026 may make an implicit misrepresentation if it states, "In *Petrella v. Metro-Goldwyn-Mayer*, the Ninth Circuit held that the doctrine of laches can bar a copyright claim brought within the statute of limitations. Petrella v. Metro-Goldwyn-Mayer, Inc., 695 F.3d 946, 956 (9th Cir. 2012)." That assertion is literally true. But to the extent the sentence implicitly presents the stated proposition as good law—as such an assertion in a brief is likely to do—it is a misrepresentation, because the Supreme Court since overruled the Ninth Circuit on this very point of law.[96]

Propositional validity can be easy to formalize if referential validity is assumed. Some legal propositions are straightforward to categorize as either correct or incorrect. As an example of a correct legal proposition, Magesh et al. offer, "The right to same sex marriage is protected under the U.S. Constitution. *Obergefell v. Hodges*, 576 U.S. 644 (2015)."[97] As an example of an incorrect legal proposition, they offer, "There is no right to same sex marriage in the United States."[98] Of course, a lawyer's skill is to "get to maybe,"[99] and many statements of law may not be provably correct or incorrect.[100] But Magesh et al. appear to limit their queries to provable or disprovable matters, such as a case's holding or the answer to a bar-exam question.[101]

Though it resembles referential validity, Magesh et al.'s concept of groundedness is better understood as propositional validity. Recall their

---

[96] Petrella v. Metro-Goldwyn-Mayer, Inc., 572 U.S. 663, 685 (2014).

[97] Magesh et al., *supra* note 42 at 221.

[98] *Id.*

[99] Richard M. Fischl & Jeremy Paul, GETTING TO MAYBE: HOW TO EXCEL ON LAW SCHOOL EXAMS (1999).

[100] Perhaps for this reason, courts in some contexts evaluate legal assertions as statements of opinion rather than statements of fact. *See, e.g.*, Coastal Abstract Serv., Inc. v. First Am. Title Ins. Co., 173 F.3d 725, 731 (9th Cir. 1999) ("Absent a clear and unambiguous ruling from a court or agency of competent jurisdiction, statements by laypersons that purport to interpret the meaning of a statute or regulation are opinion statements, and not statements of fact."). *But see* Azurity Pharms., Inc. v. Edge Pharma, LLC, 45 F.4th 479, 497 (1st Cir. 2022) ("a law or regulation could be 'so clear on its face that no good faith doubt concerning its interpretation would be possible . . .' such that it is 'so clear as to be a fact for Lanham Act purposes'" (quoting Dial A Car, Inc. v. Transportation, Inc., 82 F.3d 484, 489 n.3 (D.C. Cir. 1996))).

[101] Magesh et al., *supra* note 42 at 223.

definition: "[a] response is *grounded* if the key factual propositions in its response make valid references to relevant legal documents."[102] The focus of this inquiry is whether a cited source supports the proposition for which it is cited, not whether a source has been cited at all — or, if an attempted citation does not unequivocally refer to a single source, what source has been cited. Instead of asking whether a legal proposition *X* is correct, groundedness asks whether a cited authority *Y* supports a proposition *X*. This is an inquiry into propositional validity. Crucially, Magesh et al. do not elaborate a process for identifying *Y*, likely because within their limited context of RAG technologies that link to supporting sources, they can anticipate that *Y* will be unambiguous. In contrast, examining legal filings in the wild — as we do — requires identifying whether a given string of text in fact constitutes a citation and, if so, what legal decision it cites. Revealingly, Magesh et al. only analyze groundedness for responses that qualify as "correct."[103] If groundedness corresponded to referential validity, it would make sense to analyze it separately from propositional validity, since a citation to a case can successfully refer to a case even if it misrepresents what that case holds. Put another way, a citation can be referentially valid yet propositionally invalid: take, for example, the assertion, "There is a right to same-sex marriage in the United States. *Roe v. Wade*, 410 U.S. 113, 116 (1973)."

C. What, Exactly, Is a "Hallucination"?

To call text a "hallucination" is to make particular claims about its provenance and its content.[104] The claim about provenance is clearer: a hallucination has to be the result of generative-AI usage. The claim about content is underspecified. At minimum, to be a "hallucinated citation," we believe the text in question must be either propositionally invalid or referentially invalid. However, unless one imposes a standard of strict referential validity—which is harsher than the standard applied to legal writing produced by other means,[105] and which is not a requirement for a citation to be intelligible—a formal definition of a hallucination is elusive,

---

[102] *Id.* at 221.

[103] *Id.* ("Groundedness is only applicable for correct responses.").

[104] *See, e.g.*, Emmy Liu et al., *Position: We Need a Unified Definition of Hallucination (It's The World Model, Stupid!)*, in PROC. 43RD INT'L CONF. ON MACH. LEARNING (2026) (reviewing numerous existing definitions of hallucination and proposing a unified definition that hallucinations are large language model outputs that reflect inaccurate internal world modeling).

[105] *See* note 70, *supra*; notes 93, 174 *infra*.

because citations may remain referentially valid even if they depart from strict correctness in unexpected ways.[106] Moreover, these inquiries are not independent: questions of provenance inform questions of validity. Interpreting text necessitates that a reader assume, or ascertain, information about how that text was produced. Text that would qualify as a referentially valid citation if written by a literate human might no longer qualify as such if all of its readers knew that it had been inscribed by an ant that fell into an inkwell and then walked haphazardly over a piece of paper.[107] When inquiries stipulate the provenance of apparent citations—as prior work on hallucinated citations does—they make questions of validity dramatically simpler than such questions are where real-world legal filings are concerned.

Relatedly, even assuming that hallucinations can be differentiated formally from other sorts of erroneous citations, can hallucinated case citations be ascertained by inspecting the possibly hallucinated text alone? We conjecture that the answer is "not always," as elaborated in the next three Subsections.

1. The Formal Content of a "Hallucination" Is Hard to Specify

Legal citations can fail by being referentially invalid or propositionally invalid, as discussed above. But a richer typology is worth propounding, because existing scholarship has not thoroughly explored how AI-generated citations can fail—and how they can succeed.

The existing literature has explored AI-generated citations and propositional validity[108]; the underexplored question is how AI use affects referential validity. Existing work assumes that AI-generated citations are valid if they satisfy strict referential validity and that they are "hallucinations" if they do not.[109] On one hand, the rule may be too strict: there are good arguments, elaborated below, that AI output can meaningfully cite sources even if it fails strict referential validity. On the other hand, perhaps the rule isn't strict enough: there are arguments that

---

[106] *See* Part I.C.2.a), *infra* (discussing the recurring nonexistent, but intelligible, citation "Chelates Corp. v. Citrate, 477 U.S. 317").

[107] *See* Hilary Putnam, *Brains in a Vat*, *in* REASON, TRUTH, AND HISTORY 1 (1981) (providing example of an ant drawing in sand what appears to be a portrait depicting Winston Churchill).

[108] *See* Part __, *supra*.

[109] *E.g.*, Liu, Stammbach & Henderson, *supra* note 6.

LLMs can *never* output meaningful legal citations. Both arguments are worth taking seriously.

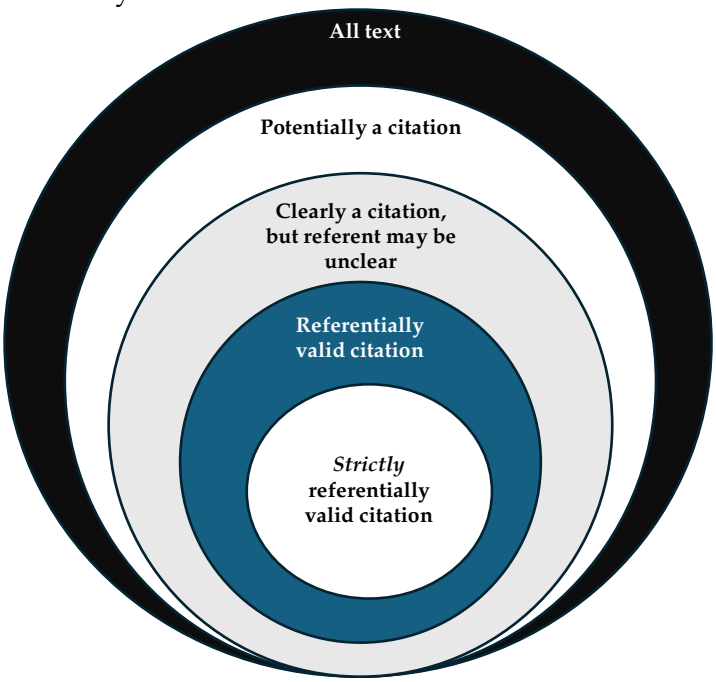


Figure 2: Taxonomy of Referential Validity

The latter view—that AI output is never meaningful—amounts to an argument that AI is *always* hallucinating.[110] This argument presents LLMs as "stochastic parrots" whose output is *per se* meaningless because it is not produced with communicative intent.[111] On this view, AI output that looks like a well-formatted citation to an existing case and appears in a context where such a citation would be semantically appropriate would nevertheless lack referential meaning because the LLM generating the output lacks communicative intent.[112]

---

[110] Cf. Chang, *supra* note __, at 14 (calling generative AI "a tool that does not know the difference between right and wrong, a tool whose architecture renders the distinction meaningless to its operation"); Matthew Sag, *AI Policies for Law Schools*, EMORY LEG. STUD. RES. PAP., 4 (2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5619534 ("[I]n some senses, LLMs are always hallucinating—it's just that some hallucinations are more useful than others.").

[111] Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* □, *in* PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 616–17 (2021), https://dl.acm.org/doi/10.1145/3442188.3445922.

[112] *Id.*

On the other hand, there are strong arguments for treating AI-generated citations as meaningful citations, even when they deviate from formal perfection. A premise of prior legal-hallucination research is that tools that consistently generated flawless legal citations would be useful sources of legal information.[113] If we grant that LLMs can successfully refer to cases when their outputs conform exactly to proper citation formats, then why can't these outputs also successfully refer to cases whenever they are intelligible as appropriate citations? It is unlikely that a reader will be confused by AI output that discusses a children's book series titled the "*Berenstein Bears*" (although the series of note is spelled *Berenstain Bears*), just as it is unlikely that a reader will find substantive fault with a mention of the New Jersey trespass case "*State v. Shacke*" (rather than correct caption, *State v. Shack*). If it is not a grave problem for a human lawyer to occasionally mistype a pincite, reporter number, or party name—so long as their citation remains propositionally and referentially valid—then is it all that worrisome if AI output occasionally does so, too? Calling an AI-generated citation a "hallucination" insinuates that it is a worthless fabrication. While some AI output certainly fits that description, output that has minor formal defects but is nevertheless readily intelligible as a correct citation probably does not. Indeed, given the prevalence of citation errors in published legal decisions, it is possible that an ostensible hallucination by a generative-AI tool could in fact be a verbatim regurgitation of an authoritative legal decision that *itself* cited a case incorrectly, quite possibly due to human error.[114]

In the 1950s, John Henry Merryman wrote of legal citations, "presumably a citation means something to the person citing, and presumably he anticipates that it will have some meaning to a reader."[115] While LLM-assisted drafting may have cast the universality of the former presumption into some doubt, legal filings do not present the hardest questions about the meaning of AI output as such, because a legal filing must necessarily be adopted by a human signatory. If the signatory actually reviews the filing and adopts the error as her own, then there is no

[113] *See* Dahl et al., *supra* note __, at 11 ("[A]ny misstatement of the binding content of [historical case] law would make an LLM quickly lose any professional or analytical utility"); Magesh et al., *supra* note __, at 9 (describing AI-generated responses as "accurate" if they "are both correct and grounded."); Patty Liu et al., *supra* note __, at 11 (asserting that a propensity to hallucinate diminishes AI's utility to litigants).

[114] *See* notes 120–130, *infra* (collecting examples).

[115] John Henry Merryman, *The Authority of Authority: What the California Supreme Court Cited in 1950*, 6 Stan. L. Rev. 613 (1954).

"stochastic parrots" problem; the text can be construed as meaningful by virtue of its review and ratification by an intentional human. In this circumstance, so-called "hallucinations" and scrivener's errors may not be all that distinct: they are both problematic only insofar as they undermine referential validity.[116] Conversely, if *unreviewed* AI-generated text has found its way into a legal filing, then the signatory has falsely represented to have reviewed it—and, in doing so, has agreed to be held responsible for its contents. On the "stochastic parrots" view, these citations lack meaning because they lack grounding in human intent—but the lack of meaning is beside the point, because the signatory has agreed to be held responsible for semantic failure.

Some AI output may be more spectacularly wrong than some human handiwork. Courts note that some errors—such as transposing a few characters—seem more likely to be the product of human mistake than, say, citing nonexistent case captions or fabricating lengthy quotations.[117] But this is an empirical assumption about the sorts of output that humans and AI tend to produce, and the provenance of an error will not always be obvious. Some invalid citations may, in context, be quite obviously explained as human mistakes; some may appear to be typos but be the product of AI; and some may bear the hallmarks of AI hallucinations yet be unrelated to AI use. These categories are challenging to differentiate, especially in an automated manner.

2. Hallucinated Case Citations Cannot Be Identified from Text Alone

Can a hallucination be identified from the suspect text alone? For three reasons—one qualitative observation based on our research experience, and two theoretical reasons—we conjecture not.

First, our experience handling our dataset illustrated just how difficult it can be to detect genuine hallucinated case citations, even for experienced human interpreters working methodically. Subsection (a), *infra*, elaborates these difficulties by way of example. Second, if a hallucination must be AI-generated, there is no foolproof method for proving that particular text was generated by AI, and programmatic AI detection requires a minimum quantum of text that is far longer than the

[116] This does assume that scrivener's errors and hallucinations affect propositional validity equally.

[117] *See* note 181, *infra*.

typical citation, as further discussed in Subsection (b), *infra*.[118] Third, if we reject the impractical standard of strict referential validity and instead require only referential validity, then formalizing the definition of hallucination requires formalizing referential validity. Formalizing when text "refers" may be an intractable linguistic task; at the very least, the determination may require extensive additional context beyond the citation in question. We leave further examination of this third point to future scholarship that we hope will involve discourse across not only computer science and law but also linguistics, philosophy, and other relevant fields.

a) "Chelates v. Citrate" and Other Stories

Consider the following example. Celotex Corp. v. Catrett, 477 U.S. 317 (1986) is a frequently cited case addressing a ubiquitous legal issue, the standard of review for summary judgment. In our dataset, we detected two occurrences of the citation "Chelates Corp. v. Citrate, 477 U.S. 317," which our system flagged as invalid citations. Because this citation looked potentially AI-generated—it had party names that seemed nonsensical and bore no apparent relation to the case located at the cited page of the U.S. Reports[119]—we singled it out for manual review. What we found illustrates the difficulty of classifying citations as "hallucinations" based on their formal characteristics.

It turns out that, between 1999 and 2021, 41 federal district court opinions in 25 different jurisdictions cited "*Chelates Corp. v. Citrate*" as if it were a real case.[120] This obvious error has the hallmarks of a hallucination, but that label is inapposite. First, if hallucination is a phenomenon limited to generative AI, then many of these erroneous citations are not hallucinations because they cannot have resulted from generative-AI usage.[121] Second, while these citations fail strict referential validity, it seems

[118] *See* Xuandong Zhao et al., *SoK: Watermarking for AI-Generated Content*, *in* PROC. OF IEEE SYMP. ON SEC. & PRIVACY (2025); Miranda Christ et al., *Undetectable Watermarks for Language Models*, *in* PROC. OF CONF. ON LEARNING THEORY (2024); *Why Does Pangram Have a Minimum Word Count?*, PANGRAM LABS, https://www.pangram.com/blog/why-does-pangram-have-a-minimum-word-count (last visited July 12, 2026).

[119] *Cf. Schlichter*, 339 Cal. Rptr. 3d at 268 ("It is difficult to understand how Grotke's four spurious citations could possibly be mere clerical errors . . . . The spurious citations do not involve the mere omission or addition or transposition of one or several digits.").

[120] According to a Westlaw search for the phrase "Chelates Corp. v. Citrate" conducted on July 12, 2026.

[121] This conclusion depends on the assumption that a federal court was unlikely to have relied upon the outputs of generative AI before, at the earliest, ChatGPT's release in late 2022.

safe to assume that they always or almost always satisfy both referential validity and propositional validity. A legally informed reader who comes across a reference to "Chelates Corp. v. Citrate, 477 U.S. 317" in a federal court's discussion of the summary-judgment standard is going to recognize that this reference points to *Celotex v. Catrett*, thus satisfying referential validity; similarly, it is reasonable to presume that the courts that cited "Chelates Corp. v. Citrate" generally did so in support of valid propositions about *Celotex*.

The etiology of the *Chelates v. Citrate* pseudo-citation is unclear. The recitation of the summary judgment standard is boilerplate that is unlikely to vary between opinions, so an initial scrivener's error may have metastasized nationwide through inattentive copying-and-pasting. Separately or additionally, given the breadth of jurisdictions that recapitulate the error, it may have resulted from a more rudimentary form of artificial intelligence: autocorrect.[122] Neither "Celotex" nor "Catrett" is a dictionary word; meanwhile, "chelates" and "citrate" both appear in the dictionary and are somewhat orthographically similar to the words they replace, which may have led multiple judges' autocorrect engines to make the substitution independently.[123] Further corroboration for the autocorrect theory is the fact that 11 federal district court opinions cite "Chelates Corp. v. Catrett" and an astonishing 366 opinions from federal district courts and the Court of Appeals for the Fifth Circuit cite "Celotex Corp. v. Citrate."[124] (Of course, the autocorrect theory and the inadvertent-copying-and-pasting theory are not mutually exclusive, and both probably help to explain the phenomenon.)

The *Chelates v. Citrate* phenomenon is far from a one-off. On a hunch, we plugged in another leading civil-procedure case involving an uncommon party name—*Ashcroft v. Iqbal*—and found that federal decisions were replete with references to the case that failed strict referential validity.

[122] Autocorrect is known to have caused errors in legal filings; for example, it appears to have caused a lawyer to file a brief that used the phrase "sea sponge" instead of "sua sponte." Robert D Lang, *Auto-Correct: Changing Sua Sponte to Sea Sponge; a Mixed Blessing for Attorneys*, 32 SYRACUSE J. SCI. TECH. L. 1, 2 (2015).

[123] *Definition of CITRATE*, (July 2, 2026), https://www.merriam-webster.com/dictionary/citrate; *Definition of CHELATES*, (July 2, 2026), https://www.merriam-webster.com/dictionary/chelates.

[124] According to Westlaw searches for the quoted citations conducted on July 12, 2026.

Among them were citations to "Igbal;"[125] "Iqubal;"[126] the dastardly "lqbal," which replaces a capital "I" with a lowercase "L;"[127] and the enigmatic "*IAshcroft v. iIbal*," which probably resulted from an incomplete mastery of the keyboard shortcut for italicizing text[128]; as well as numerous cases that reproduce the reporter citation to *Iqbal* while listing the caption of a closely related case, *Bell Atlantic Corporation v. Twombly*,[129] including several that cite the fictive *Ashcroft v. Twombly*.[130]

These phenomena illustrate several points. First, they demonstrate that real-world citations may be referentially valid despite failing strict referential validity and despite containing an error more serious than a simple mistyped or transposed character. Second, they demonstrate that pre-2022 human errors may sometimes resemble post-2022 AI hallucinations. And third, they corroborate our conjecture that there is no principled way to distinguish this kind of citation error from a hallucinated citation simply from inspecting the offending text.

b) Programmatic Hallucination Detection Is Likely Impossible

We justify our conjecture in two parts as follows. First, we conjecture that any reasonable definition of "hallucination" will include text strings[131] that could also plausibly be produced by humans ("the preliminary conjecture").[132] Second, if the preliminary conjecture is true, and if human-authored text is not considered within the scope of "hallucination," then

[125] *See, e.g.*, Sheeley v. McCormack, No. 2:15-CV-38-FTM-29DNF, 2015 WL 1456652, at *1 (M.D. Fla. Mar. 30, 2015)

[126] *See, e.g.*, In re McCallan, 599 B.R. 361, 368 (Bankr. M.D. Ala. 2019); Halvorson v. Othman, No. 5:12-CV-06775, 2013 WL 4545737, at *2 n.3 (S.D. W. Va. Aug. 28, 2013).

[127] *See, e.g.*, Whelden v. U.S. Bank Nat'l Ass'n, 494 F. Supp. 3d 68, 73 (D. Mass. 2020); Walker v. United States, 117 Fed. Cl. 304, 315 (2014).

[128] Savarese v. City of New York, 547 F. Supp. 3d 305, 353 (S.D.N.Y. 2021).

[129] In re Key Energy Servs., Inc. Sec. Litig., 166 F. Supp. 3d 822, 828 (S.D. Tex. 2016); Bar Grp., LLC v. Bus. Intel. Advisors, Inc., 215 F. Supp. 3d 524, 540 (S.D. Tex. 2017); Lowman v. Nevada, No. 3:12-CV-00592-MMD, 2013 WL 2096482, at *1 (D. Nev. May 14, 2013). In Westlaw's version of these cases, the citation links to *Iqbal*.

[130] West v. KHI Sols., Inc., No. 1:23-CV-00830-SEB-MKK, 2024 WL 2188547, at *2 (S.D. Ind. May 15, 2024); Jacob v. Pillar Ins. Co., No. 1:14-CV-012, 2015 WL 12910737, at *1-2 n.2 (E.D. Tex. Dec. 21, 2015).

[131] In this Section, we adopt the computer-science term "string" or "text string," which refers to any sequence of characters representing text.

[132] We further conjecture that any reasonable definition of "hallucination" will include text strings *S* that are reasonably likely to be generated by both AI and by humans. This does not necessarily mean that the likelihood of generation of any such string *S* by AI and by humans is similar, but simply that both likelihoods are non-negligible.

programmatic hallucination detection based on just the suspect text is impossible ("the impossibility statement").

The preliminary conjecture is socio-technical and cannot be rigorously proven in a technical sense, as it depends on notions of reasonableness and plausibility which are not and cannot be technically well defined. The impossibility statement, however, is a purely technical statement that admits a simple logical proof, as follows.

Consider any reasonable definition of "hallucination" as some kind of AI-generated content. To show the impossibility, we may assume the first and second conjectures: that is, the present definition of "hallucination" includes some text strings that could also plausibly be produced by humans, and considers those text strings to be hallucinations only if they are AI-generated. Now consider any such text string *S*, and suppose *S* is input to any candidate hallucination-detection algorithm ***A*** which classifies its inputs into two categories "hallucination" and "not hallucination." The input sequence of characters is identical regardless of whether it was produced by AI or a human, so it is impossible for ***A*** to distinguish between these cases — yet that is exactly the criterion that determines whether or not *S* is a hallucination according to the definition under consideration. Therefore, ***A*** must sometimes be wrong on input strings that could plausibly be produced by AI or humans.

The logical argument in the preceding paragraph suffices on its own to support the impossibility statement. We further bolster our argument with the result, recently established in scholarship on AI watermarking and AI detection within computer science, that robust programmatic AI detection is impossible for text strings below a minimum length (or, to be more precise, for distributions of text strings that have less than a minimum amount of entropy)[133] and typical citations fall far short of this minimum threshold.[134] The intuition behind these technical results in the computer-science literature are, in part, related to the reasoning we have given in the preceding paragraph.[135]

Having established the impossibility statement, let us return to the preliminary conjecture. While it does not admit a similarly rigorous proof, we believe in the first conjecture based on our cumulative research

[133] Entropy is an information-theoretic concept that, roughly speaking, measures the amount of uncertainty in an output distribution. *See generally* Henry Pinkard & Laura Waller, A Visual Introduction to Information Theory (Mar. 6, 2026), available at arXiv: https://arxiv.org/abs/2206.07867.

[134] *See* Zhao et al.; Christ et al., *supra* note __.

[135] *Id*. Though the intuition is similar, the technical details are more involved. *See id.*

experience with examples of erroneous citations in our dataset, as discussed in Subsection (a), *supra*: in a nutshell, humans make a wide range of (often simple) mistakes, and AI tools may generate the same mistakes either by chance or because those human errors (or other similar human errors) are in their training data. An additional reason to believe in the first conjecture is that now that LLMs are a widely known technology that has drawn substantial public attention, humans may refer to or imitate AI outputs in a range of contexts: for example, in parody, in discussing prompt engineering, in discourse about AI usage, or in deliberate quotation of AI hallucinations.

If our conjecture is correct, it means that for any candidate hallucination-detection algorithm that simply takes the suspect text as input, there will be inputs on which it is wrong. Moreover, if our strengthened preliminary conjecture[136] holds, then some inputs on which the algorithm is wrong will consist of text that is reasonably likely to be produced both by AI and by humans. These insights may suggest that future hallucination-detection methods will need to be much more complex and context-dependent if they are to work reliably.

That said, our conjecture does *not* rule out the possibility of an algorithm that takes just the suspect text as input and correctly detects hallucinations for some inputs, or even for most inputs. At the same time, the possibility also remains that close-to-perfect accuracy will be unattainable in practice. Tools that are accurate on some or most inputs may still be useful in a range of contexts, such as to flag suspect citations for further human inspection. However, such tools are inappropriate for use in other important contexts, including for triggering penalties imposed upon humans for improper AI use, or for empirical research to study the prevalence of hallucinated content at scale (such as ours).[137]

Nor does our conjecture rule out the possibility that humans may have highly accurate intuitions about whether a particular erroneous citation is AI-generated or not. Our own research experience working with our dataset of court filings suggests that in some cases, humans may have stronger and more accurate intuition than the best hallucinated-case-detection algorithms to date. At the same time, our own human intuition failed to lead us to guess the provenance of *Chelates v. Citrate* without additional research into the surrounding context, and it seems likely to us

[136] *See supra* note __.

[137] *See infra* Part III.C (discussing distinct use cases of citation-verifying technology that require distinct performance characteristics).

that there will be many situations like that one, where substantial additional context is required to understand the source of an error.[138]

We conclude with a note of caution, highlighting the limitations of AI-detection tools, which have been a subject of intense interest far beyond hallucinated citations in recent years. For example, detecting AI-generated content has been considered of interest for educational contexts where AI use is prohibited, for tracking and mitigating the spread of misinformation and "deepfakes," and more. Yet public and academic commentary to date indicates that current detection tools are highly fallible, and their problems cannot be fully mitigated with known techniques. AI-detection tools are prone to classifying AI-generated text as human-authored as well as biased in troubling ways, such as against non-native language speakers.[139] We urge a cautious and skeptical approach when deploying these tools and advise against reliance upon them, especially in high-stakes environments, including courtroom contexts.

II. The Doctrinal Stakes of Citation Validity

Classifying a legal citation as an AI hallucination has serious practical consequences. Submitting filings with hallucinated citations or other hallucinated content may amount to a violation of procedural law; attorneys submitting such filings may, moreover, violate rules of professional conduct. Attorneys and litigants may be sanctioned for any of the above. Hallucinated citations have spawned a sanctions jurisprudence unto themselves. As of May 21, 2026, a database maintained by Damien Charlotin listed 1,458 cases worldwide in which "the court or tribunal has explicitly found (or implied) that a party relied on hallucinated content or material."[140] Over 1,000 of those cases are from the United States; the database reports at least 85 awards of monetary sanctions from that subset.[141]

---

[138] *See supra* Subsection (a).

[139] Marijke Van Vlasselaer, Filip Van Droogenbroeck & Bram Spruyt, *Who Wrote This? Evaluating the Reliability of AI Detection Tools in Higher Education*, 22 INT'L J. EDUC. INTEGRITY 16 (2026); *AI-Detectors Biased Against Non-Native English Writers,* STAN. U. HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (May 15, 2023), https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers.

[140] Charlotin, *supra* note 9. *See also* Hall v. Acad. Charter Sch., No. 2:24-CV-08630-JMW, 2025 WL 2256653, at *5 (E.D.N.Y. Aug. 7, 2025) (collecting cases); United States v. Hayes, 763 F. Supp. 3d 1054, 1071–72 (E.D. Cal. 2025), reconsideration denied, No. 2:24-CR-0280-DJC, 2025 WL 1067323 (E.D. Cal. Apr. 9, 2025) (same).

[141] Charlotin, *supra* note 9.

A. Professional Conduct: Competence, Supervision, and Candor

Most federal and state courts base their professional conduct rules closely on the American Bar Association (ABA)'s Model Rules of Professional Conduct.[142] There are several Model Rules that may apply to attorneys' use of hallucinated citations, which we describe next.

Model Rule 1.1 (Competence) requires attorneys to "provide competent representation,"[143] and Comment 8 to Rule 1.1 further notes that a competent lawyer "should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."[144] The ABA and many state bar associations have recently issued ethics guidance for attorneys' use of generative AI tools, which consistently emphasize competence as a core professional duty that careless use of AI, including reliance on hallucinated information, could violate.[145]

Model Rules 5.1 (Responsibilities of Supervisory Lawyers) and 5.3 (Responsibilities Regarding Nonlawyer Assistance) require lawyers to "make reasonable efforts to ensure that" the conduct of anyone under their direct supervision complies with the Model Rules, and holds the supervising lawyer "responsible" for the subordinate's professional misconduct under certain circumstances.[146] One common response of lawyers to the revelation of hallucinated citations in their filings is to claim that the work was initially performed by a junior associated, intern, or paralegal.[147] Model Rules 5.1 and 5.3 make clear that the supervising attorney should, in such cases, make reasonable efforts to ensure the

---

[142] Nathan M. Crystal & Grace M. Giesel, PROFESSIONAL RESPONSIBILITY: PROBLEMS OF PRACTICE AND THE PROFESSION 14–15 (8th ed. 2024)

[143] MODEL RULES OF PRO.CONDUCT r. 1.1 (AM. BAR ASS'N 2025).

[144] *Id.* at cmt. 8 (AM. BAR ASS'N 2025). Comment 8 was amended in 2012 to add the language about technology; almost all states have since adopted some version of Comment 8's duty of technological competence. Am. Bar Ass'n, *Rule 1.1, Comment [8] technological competence*, https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/mrpc1-1-comment-8.pdf (last updated Apr. 4, 2023); Esquire Deposition Solutions, District of Columbia Adopts Duty of Technology Competence (Apr. 22, 2025) https://www.esquiresolutions.com/district-of-columbia-adopts-duty-of-technology-competence.

[145] ABA Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024); *e.g.*, Cal. Bar Assoc. Comm. on Pro. Resp. & Conduct, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law (2023); N.Y. Bar Assoc. Comm. on Pro. Ethics, Advisory Op. 2024-5 (2024).

[146] MODEL RULES OF PRO.CONDUCT r. 5.1, 5.3 (AM. BAR ASS'N 2025).

[147] *See* Charlotin, *supra* note 9.

subordinates do not improperly rely on AI tools, and to check the subordinates' work. As documented in Charlotin's database, the supervising lawyers are often still held responsible in such situations. Moreover, a 2012 amendment to Model Rule 5.3 changed the word "Assistants" to "Assistance" and amended Comment 3 to clarify that the rule applies to lawyers' use of technology and technology services as well as human assistants; therefore, attorneys likewise have a duty to make reasonable efforts to ensure that any technology (services) used also comply with their professional obligations.[148]

Model Rule 3.1 (Meritorious Claims) requires attorneys not to "bring or defend a proceeding, or assert or controvert an issue therein, unless there is a basis in law and fact for doing so that is not frivolous."[149] Model Rules 3.3 (Candor Toward the Tribunal) and 8.4 (Misconduct) prohibit attorneys from "mak[ing] a false statement of fact or law to a tribunal"[150] and from "engag[ing] in conduct involving dishonesty, fraud, deceit or misrepresentation" or "conduct that is prejudicial to the administration of justice."[151] The submission of hallucinated material to courts could qualify as any or all of these types of proscribed conduct.

Courts have treated attorneys' reliance on hallucinated information to violate state or federal counterparts of all of the above Model Rules of Professional Conduct, sometimes all simultaneously.[152]

B. Federal Rules of Civil Procedure: Rule 11(b)

In federal court, the rule that polices hallucinated citations is Rule 11 of the Federal Rules of Civil Procedure. Rule 11 provides that by presenting a legal filing to the court, an "attorney or unrepresented party certifies that to the best of the person's knowledge, information, and belief, formed after an inquiry reasonable under the circumstances" that, among other things,

---

[148] AM. BAR ASS'N HOUSE OF DELEGATES, RESOLUTION 112 (2019), at 6 ("In 2012, the title of Model Rule 5.3 was changed from 'Responsibilities Regarding Nonlawyer Assistants' to 'Responsibilities Regarding Nonlawyer Assistance.' The change clarified that the scope of Rule 5.3 encompasses nonlawyers whether human or not."), https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf; MODEL RULES OF PRO.CONDUCT r. 5.3 at cmt. 3 (AM. BAR ASS'N 2025).

[149] MODEL RULES OF PRO.CONDUCT r. 3.1 (AM. BAR ASS'N 2025).

[150] MODEL RULES OF PRO.CONDUCT r. 3.3 (AM. BAR ASS'N 2025).

[151] MODEL RULES OF PRO.CONDUCT r. 8.4 (AM. BAR ASS'N 2025).

[152] *E.g.*, State v. Coleman, 280 N.E. 3d. 1042, 1064– 67 (Ohio Ct. App. 2026); Lnu v. Blanche, 177 F.4th 1014, 1026–27 (9th Cir. 2026); In re Baby Boy, 271 N.E. 3d 524, 552 (Ill. App. Ct 2025); United States v. Hayes, 763 F. Supp. 3d 1054, 1064 (E.D. Cal. 2025).

"the claims, defenses, and other legal contentions are warranted by existing law . . . ."[153] A court can impose deterrent sanctions—including financial penalties, referrals to the bar, and orders directing particular behavior[154]—on a party who violates Rule 11.[155] However, "[a]ny Rule 11 sanction should be 'made with restraint'" and "[s]anctions should not be imposed 'for minor, inconsequential violations of the standards prescribed by subdivision (b).'"[156]

The precise basis for sanctioning parties who file hallucination-ridden documents is worth spelling out. Rule 11 requires a party to certify to have conducted "an inquiry reasonable under the circumstances" establishing that, among other things, the arguments in a given submission are "legally tenable" and "warranted by existing law."[157] At a minimum, this standard "require[s] that attorneys read, and thereby confirm the existence and validity of, the legal authorities on which they rely."[158] When an attorney submitted a brief to the Second Circuit that cited "non-existent" cases produced by ChatGPT, the court determined that she "made no inquiry, much less the reasonable inquiry required by Rule 11 and long-standing precedent, into the validity of the arguments she presented" and referred her to the court's grievance panel.[159] This analysis treats a hallucinated citation as grounds for inferring a lack of due care: the sanctionable conduct is submitting a filing without confirming the validity of the legal assertions it contains.

Many prominent cases involving generative AI misuse invoke Rule 11. As an early and influential case from the Southern District of New York, *Mata v. Avianca*, explained, hallucinated citations may violate Rule 11 because "[a] fake opinion is not 'existing law' and citation to a fake opinion

---

[153] Fed. R. Civ. P. 11.

[154] *See, e.g.*, Total Television Ent. Corp. v. Chestnut Hill Vill. Assocs., 145 F.R.D. 375, 385 (E.D. Pa. 1992) ("Appropriate sanctions other than a monetary award may include reference to a bar association grievance committee, compulsory legal education, an oral or written reprimand, an order barring the attorney from appearing for a period of time, or even ordering the attorney who violated the rule to circulate in his or her firm a copy of the opinion in which the pleadings, motion, or other paper were criticized.").

[155] Fed. R. Civ. P. 11(c)(4).

[156] Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 462 (S.D.N.Y. 2023) (first quoting Storey v. Cello Holdings, L.L.C., 347 F.3d 370, 387 (2d Cir. 2003), then quoting Rule 11, advisory committee's note to 1993 amendment).

[157] Fed. R. Civ. P. 11(b); Cooter & Gell v. Hartmarx Corp., 496 U.S. 384, 393 (1990); Park v. Kim, 91 F.4th 610, 615 (2d Cir. 2024).

[158] Park v. Kim, 91 F.4th 610, 615 (2d Cir. 2024).

[159] Park v. Kim, 91 F.4th 610, 615-16 (2d Cir. 2024) (quoting Cooter & Gell, 496 U.S. at 393) (emphasis in original).

does not provide a non-frivolous ground for extending, modifying, or reversing existing law, or for establishing new law. An attempt to persuade a court or oppose an adversary by relying on fake opinions is an abuse of the adversary system."[160] Some courts appear to regard a hallucinated citation as a *per se* failure-to-review violation of Rule 11. An opinion from the District of Minnesota, for example, states, "Every filing in a federal court that contains citations to phony case law amounts to a violation of Rule 11(b) of the Federal Rules of Civil Procedure."[161,162] Relatedly, Cheng-chi Chang proposes reckless misrepresentation liability for attorneys who "use[] an AI tool in a legal research context where hallucination risk is documented and publicly known, perform[] no independent verification of citations, and submit[] false citations to a court."[163]

If *per se* rules like "[e]very filing . . . that contains citations to phony case law amounts to a violation of Rule 11(b)"[164] take root, it will be important to establish what exactly counts as a "citation[] to phony case law."[165] At present, the concept is underspecified. Chang's proposal, for instance, never defines what a "false citation[]" is, or precisely what sorts of documented outputs do and do not establish notice of "hallucination risk."[166] Moreover, Chang's proposal oscillates between propositional and referential invalidity without marking the distinction. Although Chang introduces the concept of legal "hallucinations" as AI-generated "case names, reporter volumes, page numbers, and quotations that read exactly

[160] Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 461 (S.D.N.Y. 2023).

[161] Turnage v. Associated Bank, N.A., No. 25-CV-3004 (ECT/DTS), 2025 WL 3052638, at *3 (D. Minn. Sept. 12, 2025), aff'd, No. 25-3092, 2025 WL 4694563 (8th Cir. Nov. 25, 2025) ("Every filing in a federal court that contains citations to phony case law amounts to a violation of Rule 11(b) of the Federal Rules of Civil Procedure.").

[162] Courts may, however, articulate a more lenient standard for *pro se* litigants on the apparent ground that a *pro se* litigant may not reasonably know or should know of a legal misrepresentation. Jones v. Kankakee Cnty. Sheriff's Dep't, 164 F.4th 967, 970–71 (7th Cir. 2026).

[163] Chang, *supra* note __ at 16.

[164] Turnage v. Associated Bank, N.A., No. 25-CV-3004 (ECT/DTS), 2025 WL 3052638, at *3 (D. Minn. Sept. 12, 2025), aff'd, No. 25-3092, 2025 WL 4694563 (8th Cir. Nov. 25, 2025) ("Every filing in a federal court that contains citations to phony case law amounts to a violation of Rule 11(b) of the Federal Rules of Civil Procedure.").

[165] Turnage, 2025 WL 3052638, at *3; *Mata*, 678 F. Supp. 3d at 461.

[166] Chang, *supra* note __ at 16. Chang does write, "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," *id.* at 7, but does not specify what "correspond[ence]" between a citation and an authentic case demands.

like real citations but correspond to nothing in any official reporter"[167]—which emphasizes referential invalidity—the statistics about hallucination rates that Chang invokes come from studies that measure propositional invalidity.[168] Chang cites hallucination measurements from studies like the Magesh et al. paper, which count AI output as a hallucination if it cites a genuine case—and thus arguably *does* "correspond to [some]thing in an[] official reporter"—but misstates that case's holding.[169]

Presumably, proponents of *per se* rules against citing "phony cases" do not propose to find wrongdoing whenever a filing fails to conform perfectly to Bluebook format, misspells a party's name in a case caption, or transposes some digits in a reporter volume number.[170] (Indeed, Chang's proposal for recklessness liability appears to be limited to the submission of "false" citations *generated by AI*.[171]) And arguably, *any* referentially valid citation—even if formally imperfect—cannot be a "citation to phony case law," because the citation successfully refers to a real case.

So what is it these proposals are setting out to police, and is it different from the sorts of errors that courts already police? In particular, does it matter—and should it matter—whether a particular mistake was introduced by AI?

C. The Jurisprudence of Typos

As common sense would suggest, and as our empirical results corroborate, legal filings frequently contain typographical or proofreading errors that obfuscate citations. Indeed, courts have had to police litigants' citation practices since long before generative AI, and the resulting body of caselaw confirms that immaterial formal defects are not sanctionable. Courts do not generally treat typos as grounds for sanctions,[172] but courts

---

[167] *Id.* at 7.

[168] *Id.* at 7–8; See, e.g., Varun Magesh et al., Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools (May 30, 2024), http://arxiv.org/abs/2405.20362.

[169] Chang, *supra* note __ at 7–8; See, e.g., Magesh et al., *supra* note 104 at 16–17.

[170] Relatedly, Berkeley Law's brand-new Artificial Intelligence Policy provides, "Citations to sources that do not exist will raise a presumption of prohibited AI use." Artificial Intelligence Policy, UC BERKELEY LAW (Summer 2026), https://www.law.berkeley.edu/academics/registrar/academic-rules/artificial-intelligence-policy/. Although a typographical error could produce what is effectively a "citation[] to [a] source[] that do[es] not exist," one imagines that Berkeley's policy does not entail that every such typo will give rise to a presumption of impermissible AI use.

[171] Chang, *supra* note __ at 16–17.

[172] Sandoval v. Novartis Pharms. Corp., No. CV 14-07478 DDP MRWX, 2015 WL 789667, at *2 (C.D. Cal. Feb. 24, 2015) ("[T]he statement was a 'scrivener's error' or typo. The Court will not impose sanctions for good faith technical errors." (citation omitted)); Holiday v.

may cite widespread typos alongside other deficiencies as relevant to sanctions.[173]

Indeed, judges are often willing to read typos cooperatively.[174] One federal judge described a lawyer's "simple scrivener's error"—transposing two digits—as "a mistake with which this court sympathizes, as typos plague all of us."[175] Another wrote, "To be clear, not every error in a motion is recklessness or more. To err is human, and minor typographical errors, even in citations, occasionally occur despite attorneys' best efforts. Likewise, some factual or legal authorities are the subject of reasonable debate, and a mere disagreement with one side's view does not necessarily mean that the view is objectively false. The insertion of bogus citations is not a mere typographical error, nor the subject of reasonable debate."[176] This doesn't mean that citation errors never have consequences[177], but it does

---

United States, No. 17-CR-01370-AJB, 2025 WL 2109077, at *8 n.4 (S.D. Cal. July 28, 2025) (citing "287-at 3" instead of "287-2 at 3" does not "rise[] to the level of warranting Rule 11 sanctions"); Alexander v. Croix Townhomes LLC, No. 222CV02051CDSEJY, 2023 WL 4052328, at *3 (D. Nev. June 15, 2023) ("Brown's initial mistake was a simple scrivener's error . . . a mistake with which this court sympathizes, as typos plague all of us."); TIW Holdings LLC v. EVO Brands LLC, No. 5:23-CV-00005-AB-MAR, 2023 WL 9420825, at *1 (C.D. Cal. Nov. 27, 2023) ("[T]ypos in some of the Amended Affirmative Defenses made them hard to decipher. The Court expects better going forward or sanctions may be imposed.").

[173] *Holiday*, 2025 WL 2109077, at *8 n.4 ("[R]ampant typographical errors can be considered in a Rule 11 sanctions analysis . . . ."); Hickey v. Scott, 738 F. Supp. 2d 55, 73 (D.D.C. 2010) (sanctioning lawyer whose "submissions . . . are filled with often-inaccurate citations to irrelevant caselaw, and include innumerable and blatant typographical and grammatical errors"); Alston v. Mississippi Dep't of Transportation, No. 4:16-CV-236-DMB-JMV, 2019 WL 3430771, at *2 (N.D. Miss. July 30, 2019), aff'd, 804 F. App'x 225 (5th Cir. 2020) (declining to impose sanctions for "an inconsequential typographical error"); Roe v. Nevada, 621 F. Supp. 2d 1039, 1061 (D. Nev. 2007) (imposing sanctions for briefs containing flaws such flaws as "incorrect legal citations and references to the record, . . . countless typos and incomplete sentences, . . . undeveloped and nonresponsive arguments").

[174] Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 n.4 (Ind. Ct. App. 2026). *Cf.* Bunce v. Visual Tech. Innovations, Inc., No. CV 23-1740, 2025 WL 4231632, at *1 n.1 (E.D. Pa. Jan. 21, 2025) ("'Behm v. Lockheed Martin Corp., 460 F.3d 860 (7th Cir. 2006)' is not a case citation, nor is there a reasonably detectable typo that may lead the Court to find the applicable case citation that supports Mr. Rajan's proposition.").

[175] Alexander v. Croix Townhomes LLC, No. 222CV02051CDSEJY, 2023 WL 4052328, at *3 (D. Nev. June 15, 2023).

[176] Johnson v. Dunn, 792 F. Supp. 3d 1241, 1263 (N.D. Ala. 2025).

[177] *See supra* note 173, *supra*. Furthermore, failing to cite *any* authority, for instance, may result in waiver of claims. *See, e.g.*, Morrison v. Hinson-Morrison, 2024 WY 96, ¶ 41, 555 P.3d 944, 958 (Wyo. 2024) ("We do not consider issues unsupported by cogent argument or citation to pertinent authority.").

mean that proposals to penalize the submission of "false" citations *per se* are arguably inconsistent with the approach that courts have previously taken to incorrect citations.

Notably, several courts that have confronted the issue have discussed whether the typos were easy to look past, or made filings "hard to decipher."[178] Some moreover discuss the likely ease of spotting and correcting the mistakes at issue, suggesting easily avoidable mistakes may indicate a greater culpability or lack of due care.[179]

D. Does It Matter Whether AI Introduced the Error?

Citation mistakes in a filing are always indicia of the filer's inadvertence. Today, however, erroneous citations may be evidence either of familiar forms of oversight, or of AI misuse. Courts are beginning to confront the question of whether the provenance of the errors makes a difference. A consensus answer has yet to emerge.

Of course, whether errors' origins matter in practice depends to some degree on whether courts can even ascertain whether particular errors owe to AI misuse or to old-fashioned incompetence, especially where those who created the errors do not admit to AI use.[180] At least some courts in the post-AI era have been willing to speculate about how particular errors arose, especially if they do not resemble standard typographical errors like mistyped characters.[181] For instance, in *Rechlicz v. Price Point Builders*, the

---

[178] *See, e.g.*, TIW Holdings, *supra* note __, at *1.

[179] *See, e.g.*, Hickey, *supra* note ___, at 73.

[180] *Compare* Mata v. Avianca, supra note ___, *with* Ex Parte Lee, 673 S.W.3d 755 (C.A. Tex. 2023) (deciding not to request a disciplinary hearing or submit report to the state bar, where an attorney filed what Westlaw describes as "unverified citations, likely generated by AI tools" much as in *Mata v. Avianca*, but did not admit to AI use).

[181] L. Off. of Shawn Shearer, P.C. v. Iowa Dist. Ct. for Fremont Cnty., 27 N.W.3d 806, 815 n.3 (Iowa 2025) ("[I]t seems most likely that the cite to section 351.1 was a typographical error intended to cite to section 352.1 . . . . This apparent typo does not support sanctions."); Gittemeier v. Liberty Mut. Pers. Ins. Co., No. 4:24CV1236 JAR, 2025 WL 2939128, at *7 (E.D. Mo. Oct. 16, 2025) ("The errors in Liberty Mutual's filing are not ones in which a few letters or numbers were passed over or shuffled. Rather, the filing includes . . . various . . . inaccuracies [that] cannot be explained by typographical or vision issues."); Schlichter v. Kennedy, 116 Cal. App. 5th 24, 32, 339 Cal. Rptr. 3d 262, 268 (2025) ("The spurious citations do not involve the mere omission or addition or transposition of one or several digits. Rather, all four spurious citations are completely different from the correct citations for the actually existing cases that have those case names. Grotke's spurious citations bear the hallmarks of hallucinated citations produced by generative AI."). *See also* Oral Argument, Malkeet LNU v. Blanche at 10:10, https://www.ca9.uscourts.gov/cases/streams-videos/archive?case=24-4790&hearingDate=2025-10-20 ("Are you telling me the truth that you didn't use AI to beef up your brief here?"); *id.* at 11:25 ("[T]his doesn't look like a

Court of Appeals of Indiana recently evaluated a brief "contain[ing] a handful of record and case citations with obvious typos, and [a citation to] one case that does not exist," but concluded that "these incorrect citations appear to us to have been the result of inadvertent typographical errors and did not hinder our review of [the party's] arguments."[182] For our part, we conjecture that it is impossible to programmatically differentiate "hallucinated" citations from scrivener's errors, but we also suspect that expert readers scrutinizing the filings in context may be able to tell the difference much more reliably.[183]

The *Rechlicz* opinion went on to confront whether it mattered if errors were the product of "inadvertent typos" or incautious use of generative AI. The court suggested that the difference mattered:

> We are mindful that the increasingly prevalent use of generative artificial intelligence (AI) has led many courts to sanction attorneys whose filings included fake case citations. But we also keep in mind that this Court has typically shown grace to practitioners whose citations—though incorrect—appeared to be the result of inadvertent typos. Accordingly, while we admonish . . . counsel for his failure to adequately proofread and verify several of his legal authorities, we deny the . . . request to impose sanctions.[184]

Other judges seem likewise inclined to treat AI-introduced errors as special.[185] A sanctions opinion by a magistrate judge treated an "initial, undisclosed use of AI products to generate the first draft of [a] brief" as a wrong in itself, in addition to the distinct "failure to cite-check" the AI-generated text.[186] At least one judge has issued a standing order that

typographical error. It doesn't look like a copy and paste job. It looks like there is a use of technology being employed that is not being checked.").

[182] Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 n.4 (Ind. Ct. App. 2026).

[183] *See* Part __, *supra*.

[184] Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 n.4 (Ind. Ct. App. 2026) (citations omitted).

[185] As does at least some scholarship. *See generally* Chang, *supra* note __.

[186] *See* Lacey v. State Farm Gen. Ins. Co., No. CV 24-5205 FMO (MAAX), 2025 WL 1363069, at *3-*4 (C.D. Cal. May 5, 2025) ("The initial, undisclosed use of AI products to generate the first draft of the brief was flat-out wrong. Even with recent advances, no reasonably competent attorney should out-source research and writing to this technology – particularly without any attempt to verify the accuracy of that material."), *cited in* Chang, *supra* note _, at 18.

categorically forbids the use of generative AI to draft filings.[187] Others have issued orders or rules requiring disclosure of AI use, limiting AI use to certain court-approved AI tools, and requiring attorneys to verify and edit AI outputs (not just for accuracy, but also to eliminate bias, offensive content, and more).[188] Admittedly, these policies might simply reflect the judges' determinations that AI use makes errors more likely, rather than that AI-related errors are different in kind.

On other occasions, however, courts have concluded that the provenance of citation errors is irrelevant. These courts have reasoned the contents of a filing can be so egregious that—regardless of how they ended up in the document—their very submission demonstrates professional misconduct.[189] Perhaps the best elaboration of this position appeared in a recent decision from the Ninth Circuit. The court sanctioned two attorneys whose papers included "multiple fabricated citations and quotations," and opined,

> It is ultimately irrelevant to the disciplinary analysis . . ., however, whether [the sanctioned attorneys] or anyone at the Firm actually used generative AI or was aware of its tendency to make mistakes. Attorneys do not need cutting-edge technology to fabricate citations and make demonstrably false and unsupported statements. And the ethical and procedural rules we apply today do not turn on the source of such error. That is, the rules are not violated at the point of research and drafting, but at the point of signing and filing. If an attorney

---

[187] Hon. Michael J. Newman, "Standing Order Governing Civil Cases," at 13-14 https://www.ohsd.uscourts.gov/sites/ohsd/files/MJN%20Standing%20Civil%20Order%208.27.25.pdf. *See also* Perlman, Andrew M. "The Legal Ethics of Generative AI." SSRN Scholarly Paper No. 4735389 at 11 n.53, Social Science Research Network, February 22, 2024, https://doi.org/10.2139/ssrn.4735389.

[188] *See, e.g.*, *Standing Orders: Courts' Nascent Governance of AI Practices | Alerts and Articles | Insights | Ballard Spahr*, https://www.ballardspahr.com/insights/alerts-and-articles/2024/02/standing-orders-courts-nascent-governance-of-ai-practices (last visited July 1, 2026); https://assets.law360news.com/2455000/2455068/judicial%20council%20model%20rules.pdf; *see also generally* https://www.ncsc.org/resources-courts/ai-state-courts. *see also* Maura R. Grossman, Paul W. Grimm & Daniel G. Brown, Is disclosure and certification of the use of generative AI really necessary?, 107 Judicature 69 (2023).

[189] Heimkes v. Fairhope Motorcoach Resort Condo. Owners Ass'n, Inc., No. 1:22-CV-448-TFM-N, 2026 WL 883120, at *10 (S.D. Ala. Mar. 31, 2026) ("[E]ven if the Court takes [counsel] at his word that he did not use AI, then the conduct falls squarely within [Alabama] Rule [of Professional Conduct] 3.3 because the citations were then deliberately represented to the Court as precedent.").

> files a brief with cases or quotations that do not exist, or completely misrepresents what a real authority stands for, it generally does not matter if he pulled the hallucination or misrepresentation from the output of an artificial intelligence tool or from his own natural intelligence.[190]

A federal court in Missouri took a similar approach. In *Gittemeier v. Liberty Mutual Personal Insurance Company*, counsel cited nonexistent cases in a filing, received a written admonition from the court, and then in a subsequent filing "cited two nonexistent cases again . . .[,] misquoted or mischaracterized multiple cases," and misrepresented the contents of a treatise.[191] The explanation offered was that "the errors were typographical and/or caused by vision impairment," but the court rejected this as "simply not credible."[192] At a subsequent sanctions hearing, the attorney who authored the error-ridden briefs "assured the Court that these errors were human and not a product of artificial intelligence (AI)"; her firm presented the court with the results of "two separate technical audits, which concluded that the citation errors were human and not AI-generated."[193] Ultimately, the court concluded from the evidence presented that "the citation and quotation defects . . . were the result of egregious human error and not AI," but it imposed sanctions nevertheless.[194]

Finally, the duty of technological competence (see supra Subsection II.A) definitionally distinguishes between errors resulting from technology misuse and other errors. Courts have specifically noted the violation of the duty of technological competence in cases of acknowledged misuse of AI tools.[195] However, technological competence is but one element of the broader duty of competence, and technological and non-technological errors alike may violate the duty of competence more broadly.

In sum, courts' approaches to AI-related errors are nascent, varied, and still evolving. One view suggests that is more blameworthy to submit erroneous citations that were produced by AI than erroneous citations that were not. Another view, however, suggests that the text of the filing is what matters, irrespective of whether the attorney "pulled the hallucination or

---

[190] Malkeet Lnu v. Blanche, 177 F.4th 1014, 1020, 1025–26 (9th Cir. 2026).

[191] Gittemeier v. Liberty Mut. Pers. Ins. Co., No. 4:24CV1236 JAR, 2025 WL 2939128, at *7 (E.D. Mo. Oct. 16, 2025).

[192] *Id.*

[193] Gittemeier v. Liberty Mut. Pers. Ins. Co., No. 4:24-CV-1236-JAR, 2025 WL 3206484, at *3 (E.D. Mo. Nov. 17, 2025).

[194] *Id.*

[195] *E.g.*, Noland v. Land of the Free, 114 Cal. App. 5th 426, n. 7 (Cal. Ct. App. 2025).

misrepresentation from the output of an artificial intelligence tool or from his own natural intelligence."[196] As the remainder of the Article elaborates, we favor the latter view from a practical policy perspective — even if as a theoretical or philosophical matter, provenance may be a factor in blameworthiness.

E. Should It Matter Whether AI Introduced the Error?

There are good arguments that a hallucinated citation in a legal filing is generally a greater ethical and professional trespass than a citation that is erroneous or unintelligible due to human error. However, the practical difficulties of differentiating hallucinations from human errors suggest that enforcers of professional discipline will be unable to enforce distinct standards for AI-generated text. Yes, it is perhaps easier to sympathize with a bedraggled but hardworking lawyer who is just too incompetent to cite cases correctly than it is with a lawyer who pastes ChatGPT output into a brief, files it, and then heads to the golf course. But the truth is that neither of these people should be practicing law. Judging a brief is, to paraphrase Wimsatt and Beardsley, "like judging a pudding or a machine. One demands that it work."[197] A lawyer should be judged by the work product he submits, and that work product should be judged by its formal qualities rather than by how it was made.[198] If a brief cites controlling law to make persuasive legal arguments that support its side of the case, then the lawyer who signs it has discharged her duty regardless of whether the brief was generated by AI or tapped out on a typewriter. If anything, it's *better* for the lawyer to spend less time and effort realizing the same work product, since the services rendered will cost the client less.[199]

Admittedly, a lawyer makes a false certification—and thus violates Rule 11—whenever she signs a brief that she has not reviewed, even if the brief happens to be flawless. But in practice, Rule 11 enforcement polices works rather than faith. Realistically detectable violations will depend on the contents of the filing, not the lawyer's state of mind when she signed it.

---

[196] Malkeet Lnu v. Blanche, 177 F.4th 1014, 1020, 1025–26 (9th Cir. 2026).

[197] W. K. Wimsatt & M. C. Beardsley, *The Intentional Fallacy*, 54 SEWANEE REV. 468, 469 (1946).

[198] *Cf.* James Grimmelmann, Benjamin L.W. Sobel, & David Stein, *Generative Misinterpretation*, 63 HARV. J. ON LEGIS. 229, 286-90, 297-99 (2026) (discussing the difference between persuasive rhetoric and legitimate adjudication).

[199] The assumption that the lawyer produces the *same* work product regardless of the method used is doing a lot of work here. In practice, the lawyer will almost certainly produce different work product through different methods, and there is a reasonable debate to be had about whether that work product would be of comparable quality both ways.

The objective of professional responsibility and sanctions is to preserve the integrity of judicial proceedings, to promote justice, and to protect consumers, rather than primarily to punish misbehaving attorneys.[200] A set of rules that lets some careless attorneys skate on good luck probably represents a better use of resources than a regime that laboriously roots out all dishonesty regardless of materiality (which, in any case, may not be realistically attainable).[201]

F. The Stakes of Erroneous Precedent in Common Law

Our discussion thus far has proceeded largely from the implicit concern that catching erroneous citations may waste courts' and litigants' resources and impede the administration of justice. We turn now to consider what may happen when erroneous citations are *not* caught.

As illustrated in the story of the non-existent but commonly cited "Chelates v. Citrate,"[202] some erroneous citations *do* slip through the court system without remark. Moreover, an erroneous citation like "Chelates v. Citrate" — which was consistently invoked instead of a standard citation for the summary judgment standard, one of the most basic elements taught in U.S. law school education and staple knowledge for essentially every judge and practicing litigator — seems likely to have no impact on the conduct or outcome of a case.

Not all erroneous citations are so harmless. We have seen hallucinated citations that include invented quotes to support a litigant's

[200] *See, e.g.*, Fed. R. Civ. P. 11(c)(4) ("A sanction imposed under this rule must be limited to what suffices <u>to deter</u> repetition of the conduct or comparable conduct by others similarly situated." (emphasis added)). *See also* Ohralik v. Ohio State Bar Ass'n, 436 U.S. 447, 460 (1978) (observing that a code of professional responsibility for lawyers implicates state interests in "protecting consumers" as well as an "especially great" interest in regulating lawyers "since lawyers are essential to the primary governmental function of administering justice, and have historically been 'officers of the courts.'" (quoting Goldfarb v. Virginia State Bar, 421 U.S. 773, 792 (1975))).

[201] *See* Model Rules of Pro. Conduct r. 8.3 at cmt. 3 ("If a lawyer were obliged to report every violation of the Rules, the failure to report any violation would itself be a professional offense. Such a requirement existed in many jurisdictions but proved to be unenforceable. This Rule limits the reporting obligation to those offenses that a self-regulating profession must vigorously endeavor to prevent.").

[202] *See supra* Subsection I.__.

case,[203] purport to stand for incorrect propositions of law,[204] and appear to support frivolous claims for which there is no basis in law.[205] If such citations were to go undetected, they could end up being invoked as authoritative holdings in genuine legal decisions[206] or affect the substantive resolution of a case, and result in serious miscarriage of justice. There are at least two distinct types of harm that correspond to our categories of propositional and referential invalidity. If a court includes a propositionally invalid citation in a decision, it has either made a substantive error of law or it has mischaracterized applicable precedent. If a court includes a referentially invalid citation in a decision—even if that citation supports a legally correct proposition—it has compromised the legitimacy of its adjudication in a way that implicates due-process concerns.[207] Any of these outcomes may cause substantive or procedural harm to litigants and other stakeholders and undermine the legitimacy of the judicial system. [208] Furthermore, such cases could create precedent based on hallucinated information to which future cases might refer — including future cases not involving any AI use at all. In sum, any invalid citation has the potential to cause harm as described above.

---

[203] *See, e.g.*, Gouldy v. Chiasson, No. 4D2025-1289, 2026 WL 2104492, at *5 (Fla. Dist. Ct. App. July 22, 2026) ("[The plaintiff] quotes a nonexistent case, Barton v. McGovern, 302 So. 3d 1042 (Fla. 2d DCA 2020), for a proposition that is consistent with her argument.").

[204] *See, e.g.*, Jarrus v. Governor of Michigan, No. 25-CV-11168, 2025 WL 3461540, at *1 (E.D. Mich. Dec. 2, 2025), reconsideration denied, No. 25-CV-11168, 2026 WL 27814 (E.D. Mich. Jan. 5, 2026) ("[T]he court identified a series of citations in Plaintiffs' brief that were false citations (with real cases, but with explanations that did not reflect the case cited).").

[205] *See, e.g.*, Kruse v. Karlen, 692 S.W.3d 43, 48 (Mo. Ct. App. 2024), reh'g and/or transfer denied (Apr. 9, 2024) (awarding damages for a frivolous appeal in which "Appellant submitted an Appellate Brief in which the overwhelming majority of the citations are not only inaccurate but entirely fictitious.").

[206] Isolated instances of this type are already known to have occurred, though thankfully they were detected and the affected decision was able to be reversed. *E.g.*, H.C. v. Contreras, 2026 WL 1493401 (Cal. Ct. App. 2026).

[207] In *HC v. Contreras*, for example, the reviewing court noted that "the trial court committed at least two clear legal errors: it "relied on a fictitious case" *and* it "misstate[d]" the requirements of a statutory provision. 2026 WL 1493401, at *3-4. Neither error is harmless: reliance on fictitious caselaw is a serious process error that compromises the legitimacy of an adjudication, while misstating a statutory test is substantive legal error. *See* Grimmelmann, Sobel, & Stein, *supra* note __ at 287–88.

[208] This kind of problem is not limited to hallucinated *citations*; hallucinated information more broadly (e.g., concerning factual information, or concerning the legal propositions for which a real case stands) may have similarly deleterious effects if it goes undetected in legal proceedings.

Even though imperfect citations were commonplace before generative AI, as discussed at length earlier in this Article, there are reasons to be more concerned about harms arising from invalid citations today than before the advent of modern generative AI and hallucinations. Erroneous citations resulting from human negligence on the part of an attorney otherwise acting in good faith — that is, not seeking to misrepresent law or facts — are comparatively unlikely to be harmful in the senses we have just defined.[209] On the other hand, an attorney's negligent use of AI has much greater potential to produce invalid citations — even, many of them per document — even if the attorney is otherwise acting in good faith; if we take the attorneys' accounts as true, this appears to be exactly what happened in many of the recent cases involving hallucinated citations.[210] (AI tools may also streamline the work of attorneys acting in bad faith to mislead.) Generative AI thus has the potential to scale up the prevalence of invalid citations in legal proceedings, and in turn, the potential harm resulting from them, in a manner qualitatively different from before AI tools were available. This scalability may, in turn, make it harder to detect all or most (harmful) invalid citations in practice.

This is deeply concerning in any system of justice, but especially in a common-law system like ours, where "[g]eneral rules, underlying principles, and finally legal doctrine . . . successively emerge[] only as precedents, accumulated through the centuries," and each "decision controls the judgments of future judgments of courts in like or analogous cases."[211] An important justification for the latter principle, called *stare decisis*, is that the law should not be "susceptible to whim or caprice,"[212] or in the Founders' words, to "avoid an arbitrary discretion in the courts."[213] To rely on erroneous precedent within a common-law system thus undermines the fundamental principle of *stare decisis* as well as the source of law itself on which the justice system is premised, and not only for a single case, but

[209] *But see* Margie Alsbrook, *Untangling Unreliable Citations*, 37 GEO. J. LEGAL ETHICS 415, 415–16 (2024) ("Even before the recent avalanche of artificial-intelligence-based legal products, the modern American justice system was arguably being destabilized by the erosion and dilution of precedent.") (discussing how copying-and-pasting or relying on another lawyer's paraphrased language without checking original sources can influence case outcomes and create precedent not accurately reflecting the prior cases it cites).

[210] *See, e.g.*, Mata v. Avianca, supra note __.

[211] Harlan F. Stone, *The Common Law in the United States*, 50 HARV. L. REV. 4, 6 (1936).

[212] William O. Douglas, *Stare Decisis*, 49 CORNELL L. REV. 735, 735–36 (1949).

[213] The Federalist No. 78, at 471 (Alexander Hamilton) (Clinton Rossiter ed., 1961).

for all future cases.[214] Indeed, forming law based on hallucinated precedents — generated in moments of negligence at best and bad faith at worst — sounds to us exactly like the sort of "whim and caprice" that it is essential to avoid.

We remain hopeful that judges, clerks, and our adversarial system of litigation will ensure that invalid legal citations will be detected reliably enough so as not to distort outcomes and precedents. Still, given the heightened potential for the proliferation of invalid citations in the presence of widely available generative AI tools, it is important for researchers and practitioners to stay alert to the risks — whether to address them if the need arises, or (hopefully) to bolster confidence in the integrity of the legal system if not. Finally, the possibility of hallucinated information influencing case outcomes raises novel questions regarding the precedential validity of such decisions — an interesting topic for future research, which we hope will remain of purely theoretical significance.

III. Reflecting on Invalid Citations and Practical Considerations

Our empirical analysis demonstrates that lawyers have long filed documents full of citations that are, in the strict sense, "invalid." But there are strong reasons to think that in many cases, those citations are in fact perfectly effective—they simply suffer from minor scrivener's errors that do not detract from their referential validity and certainly fall far short of sanctionable misconduct. A few helpful interventions could ameliorate problems of defective citation, whether introduced by AI or not. The following Subsections reflect on some sources of difficulty, possible impacts and mitigations, and who should bear the burden of invalid citations.

A. Access to Justice

Although caselaw is in the public domain[215], practicing lawyers often rely on proprietary databases to access it. These databases are costly, which means that litigants with more limited resources encounter limitations on their access to public documents that well-resourced litigants do not. The problem of proprietary legal informatics long predates digitization. At least

[214] That said, reference to precedents can be important in non-common-law systems too. *See generally* C. Sumner Lobinger, *Precedent in Past and Present Legal Systems*, 44 MICH. L. REV. 955 (1946); Sebastian Lewis, *Precedent and the Rule of Law*, 41 OXFORD J. LEGAL STUD. 873 (2021).

[215] Georgia v. Public.Resource.Org, Inc., 590 U.S. 255, 264–66 (2020).

as far back as 30 years ago, law librarians have advocated for public-domain citation reform.[216]

The cost of legal databases indisputably contributes to the incidence of hallucinated citations in court filings, although hallucinations afflict well-heeled lawyers, too.[217] One attorney, whom the Southern District of New York recently sanctioned for "repeatedly fil[ing] submissions with false citations," explained that he was "unable to verify certain citations" after he "discontinued my subscription [to Casetext] due to the prohibitive cost of maintaining access" after Casetext was acquired by Thomson Reuters and absorbed into Westlaw.[218] Similarly, an attorney sanctioned in the *Mata* case explained that "he turned to ChatGPT for legal research" in part because his firm "uses a legal research service called Fastcase and does not maintain Westlaw or LexisNexis accounts. . . . [T]he firm's Fastcase account had limited access to federal cases."[219]

Considering the disadvantages realistically faced by modern-day attorneys who lack access to proprietary legal research tools, some commentators have gone as far as to suggest that *not* using proprietary technology for legal research could conceivably amount to a breach of professional responsibility.[220] While these subscriptions may be a drop in the bucket for many well-resourced firms, this is a troubling proposition for access to justice when considering the well known scarcity of affordable legal representation.[221] The real problem here lies not in the professional conduct regime, but in the reality that state-of-the-art legal research tools are costly to the point of exclusionary.

---

[216] *See, e.g.*, *Board Adopts Resolution on Citations and Public Database*, 26 AM. ASS'N L. LIBRS. NEWSL. 161 (1994) ("[T]he Executive Board of [the American Association of Law Libraries] on November 6 adopted and transmitted via the Internet a resolution strongly supporting the concepts of public domain citation formats . . . .").

[217] Nerkar, *supra* note 5.

[218] Flycatcher Corp. v. Affable Ave. LLC, No. 24 CIV. 9429 (KPF), 2026 WL 306683, at *1–2 (S.D.N.Y. Feb. 5, 2026)

[219] Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 456 (S.D.N.Y. 2023).

[220] *E.g.*, Ky. Bar Assoc. Ethics Op. KBA E-457 (2024) ("[A]s AI tools become more refined . . . not using an available AI tool may constitute a failure to meet the lawyer's duty of . . . competence under Rule 1.1."); *see also* Cass v. 1410088 Ontario, Inc., 2018 CanLII 6959 (Can. Ont. S.C. 2018) (questioning why an attorney had not used AI tools for research); Tammy Pettinato Oltz, *Educating Robot-Proof Attorneys*, 97 N. DAKOTA L. REV. 186, 189–91 (2022) (discussing a duty to use "standard methods of legal research commonly accepted in the profession" that "almost certainly includes using online sources, whether free or fee-based.").

[221] *See* LEGAL SERVS. CORP., EXECUTIVE SUMMARY: THE JUSTICE GAP 1 (2022), https://justicegap.lsc.gov/resource/executive-summary.

Public-domain legal materials are inexcusably difficult for litigants to access. Patty Liu et al. contend, "expanding public, freely[ ]available access to legal data (including all case law) would improve both human and automated citation verification performance for under-resourced filers."[222] We agree. A policy initiative like the pending Open Courts Act of 2026—which would mandate the establishment of "a centralized system with a separate public interface" for accessing federal court dockets, opinions, and orders—would represent important progress on this issue.[223]

Some commentators have seen AI as a boon for access to justice, both because it may streamline legal work by attorneys—in theory, enabling them to provide quality services to more clients at lower cost—and because it may increase access to legal knowledge by *pro se* litigants and the public at large.[224] While this is certainly an attractive vision — and while also acknowledging that some efforts to prohibit the use of AI tools for legal advice may have more to do with professional gatekeeping than with a genuine interest in consumer protection[225] — given the current and foreseeable fallibility of AI tools, we harbor serious concerns in the other direction. Screening for hallucinated citations (and other hallucinated information) threatens to consume increasing resources and slow down already overloaded judicial dockets.[226] Undetected hallucinations could lead to wrongly decided cases and tainted precedent.[227] And the public, *pro se* litigants, and represented litigants alike may be misled by these AI tools into a false sense of confidence, based on which they may take actions with serious impacts on their rights and welfare. Indeed, it is already common for laypeople to ask LLMs for "legal advice" and there has been a noticeable

---

[222] Liu, Stammbach, and Henderson, *supra* note 3 at 11.

[223] Open Courts Act of 2026.

[224] *See, e.g.*, LEGAL SERVS. CORP., REPORT OF THE SUMMIT ON THE USE OF TECHNOLOGY TO EXPAND ACCESS TO JUSTICE 10 (2013), https://www.lsc.gov/our-impact/publications/other-publications-and-reports/report-summit-use-technology-expand-access-justice; John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 FORDHAM L. REV. 3041 (2014); 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, 67–68 (2023).

[225] *See generally* Deborah L. Rhode & Lucy Buford Ricca, *Protecting the Profession or the Public? Rethinking Unauthorized-Practice Enforcement*, 82 FORDHAM L. REV. 2587 (2014); Avery et al., *supra* note __, at 71–72.

[226] *See* Merritt McAlister, Adalberto Jordan & Kimberley J. Mueller, *What Can Be Done About Backlogs*, 107 JUDICATURE 50 (2023); Evan Ochsner, *AI-Faked Cases Become Core Issue Irritating Overworked Judges*, BLOOMBERG LAW (Dec. 29, 2025).

[227] *See supra* Part II.F.

uptick in *pro se* litigants since generative AI.[228] While LLM advice sometimes serves them well, there are serious risks alongside.[229]

B. Streamlining Citation Conventions

Other problems stem from citation conventions. Some of these challenges arise from the same access-to-justice issues described in the preceding subsection. Proprietary legal databases make citation verification more difficult—recall that we were unable to verify a substantial portion of the citations that we detected, because they pointed to Westlaw or Lexis.[230] The Open Courts Act of 2026 includes a provision that would help break proprietary citation schemes. The bill provides that the system for searching and accessing court records "shall create a machine-readable neutral citation for all orders, opinions, and decisions made available through the system," consisting of the "year of the filing of the document, an abbreviated court name, and a serial number for the document."[231] Individual courts can also help address this problem by mandating the use of public-domain citation formats in their local rules. The *Bluebook*'s hyper-technical prescriptions probably aren't doing us any favors, either.

Beyond improvements to the text we use to express citations, it is possible to imagine improving the technology that lawyers use to communicate citations. Litigants filing digital documents could, for example, incorporate hyperlinks in their citations that point to the cited materials. Or after-the-fact parsing software could insert these links automatically, although it would encounter the same difficulties in identifying and resolving citations that we had to contend with in our empirical work.

Finally, open technological initiatives to better organize citations may provide essential infrastructure for both communicating citations (as discussed above) and validating citations (see infra Part III.C). An emerging literature on automated detection of hallucinated citations to scholarly publications offers an interesting point of comparison. Works in this line

---

[228] *See* Patricia Clarke, *Meet the vibe lawyers: the AI chatbots wreaking havoc in the justice system*, OBSERVER (Feb. 24, 2026); Or Cohen-Sasson, *Artificial Access to Justice: AI and the Surge in Pro Se Litigation* (Jan. 1, 2026), available at SSRN: https://ssrn.com/abstract=6864398.

[229] *See* Clarke, *supra* note __; Eike Schneiders, Joshua Krook, & Tina Seabrooke, *People trust legal advice generated by ChatGPT more than a lawyer – new study*, CONVERSATION (Apr. 28, 2025); Mike Scarcella, *AI ruling prompts warnings from US lawyers: Your chats could be used against you*, REUTERS (Apr. 15, 2026).

[230] *See* Part __, *supra*.

[231] Open Courts Act of 2026 § (b)(11), S. 4667, 119th Cong. (introduced in Senate, June 2, 2026).

avoid many of the technical challenges that we faced in our automated detection of hallucinated caselaw, thanks to the ready availability of multiple well populated scholarly databases to check against.[232]

C. Mitigation Strategies

Lawyers and legal institutions may need to adopt citation-validation tools. Indeed, such tools are already being marketed to lawyers.[233] Existing tools are imperfect,[234] but may still be valuable for some uses. Often, they "impose restrictive usage limits on free tiers or require substantial subscription fees,"[235] raising important questions about who will have access to these tools in practice; broad, affordable access will be important to bring societal benefit without exacerbating existing inequities.[236]

Potential use cases vary considerably. An attorney, for example, faces two distinct applications for a citation-checking tool: (1) precautionary use for proofreading her own documents, and (2) offensive use to find defects in filings by opposing counsel (or even, regrettably, by the court itself).[237] The performance characteristics of an ideal tool may vary with the task. A citation-checking tool designed for proofreading assistance should be ruthlessly pedantic: it should flag every invalid citation from the most footling typo to the most flagrant hallucination. Eliminating every scrivener's error may be impossible, and the occasional good-faith mistake should not trigger professional discipline, but lawyers should strive to

---

[232] *E.g.*, Diletta Abbonato, *CheckIfExist: Detecting Citation Hallucinations in the Era of AI-Generated Content* (Jan. 27, 2026), available at arXiv: https://arxiv.org/abs/2602.15871 (presenting a new "an open-source web-based tool designed to provide immediate verification of bibliographic references through multi-source validation against CrossRef, Semantic Scholar, and OpenAlex scholarly databases"); Zuyao Xu et al., *GhostCite: A Large-Scale Analysis of Citation Validity in the Age of Large Language Models* (May 14, 2026), available at arXiv: https://arxiv.org/abs/2602.06718; Yusuke Sakai, Hidetaka Kamigaito & Taro Watanabe, *HalluCitation Matters: Revealing the Impact of Hallucinated References with 300 Hallucinated Papers in ACL Conferences* (Jan. 26, 2026), available at arXiv: https://arxiv.org/abs/2601.18724.

[233] LawDroid Ltd, *CiteCheck AI - Legal Citation Verification Tool*, CITECHECK AI, https://citecheck.ai/ (last visited June 30, 2026); Ross Guberman, *Spotting Fake Case Law with Citation Validation Engines*, (Apr. 20, 2026), https://www.briefcatch.com/blog/blog-citation-validation-engines-fake-case-law; Jacqueline Schafer, *How the Largest Global Firms Are Operationalizing Responsible AI with Citation-Checking Policies*, CLEARBRIEF (June 7, 2026), https://clearbrief.com/blog/responsible-ai-citation-checking-policies.

[234] *See generally* Patty Liu et al., supra note __.

[235] Abbonato, supra note __, at 1.

[236] *See also infra* Part III.D.

[237] *See* note 5, *supra*. *See also* H. C. v. Contreras, No. F089316, 2026 WL 1493401, at *1 (Cal. Ct. App. May 28, 2026)

submit perfect briefs. Furthermore, a tool for use at the proofreading stage should facilitate not just error detection, but error correction; it could do so by offering a list of the nearest valid citations to any invalid citation it flags.

By contrast, a lawyer running hallucination-detection software on opposing counsel's filings does not have a proofreader's objectives.[238] It would be a waste of everyone's time for an attorney to bring every "invalid" citation to the attention of opposing counsel and the court; that is line-editing, not lawyering.[239] Software used in this situation should optimize for intelligibility rather than pedantry: it should help a good-faith reader identify the most probable meanings of a formally defective citation, while simultaneously highlighting citations that genuinely suggest AI misuse. Software like this needn't be a hallucination detector specifically so much as a general user-interface improvement for legal documents—it could turn a legal filing into hypertext by collating all of the cited cases, in much the same way that Westlaw and Lexis enhance the readability of the documents in their databases by turning citations into hyperlinks to the cited sources.[240]

Similarly, institutional use of hallucination-detecting software needs to be tailored to a particular objective. Even setting aside hallucination detection, courts would certainly benefit from technology that collated referenced materials and made them more seamless to access. (Indeed, courts already have rules in place that require litigants to furnish materials to them in appendices.[241]) But institutional use of citation-validation software *as such* will have to calibrate the extent to which it is to serve as a resource-saving measure versus a measure to investigate and punish AI misuse by litigants. Should the software and its interface be calibrated charitably to reconstruct the probable intent behind formally invalid citations, or should it be constructed to root out as many hallucinations as possible in order to discipline the parties responsible? In which contexts are these different configurations appropriate or useful?

---

[238] Schafer, *supra* note 141 at Cf. ("And it can perform this analysis on a filed PDF by your opponent as well.").

[239] *Cf.* Rechlicz v. Price Point Builders, LLC, 276 N.E.3d 132, 145 n.4 (Ind. Ct. App. 2026).

[240] At least one commercial software suite advertises such functionality. CLEARBRIEF, https://clearbrief.com/.

[241] *See, e.g.*, U.S. SUP. CT. R. 26(1) (specifying rules for the filing and contents of a joint appendix).

D. What's the Point of Citations? Do They Matter In Practice?

Legal citations "communicate the precedents that underline our arguments and analysis"[242] and may be seen as "the fundamental justification" for legal arguments and decisions.[243] At a basic level, litigants and courts should demonstrate through their citations that their reasoning is grounded in law rather than arbitrary assertion. In common-law systems incorporating *stare decisis* where filings and opinions are published,[244] citations delineate a chain of precedent, allowing future judges, litigants, and the public to trace how legal principles have cohered and evolved over time. Scholarly analyses of legal citations in court proceedings have found that citation practices among judges and among lawyers differ.[245] The primary scholarly theories around the practice of citation in judicial opinions include that they may indicate the true substantive basis of reasoning, or they may simply serve to legitimate (and constrain, to some extent) reasoning that was not in fact based on the cited sources; some have argued that the reality involves a little of both.[246] In litigants' filings, prior scholarship indicates that citation use serves an important persuasive role,[247] and may serve not only to assert a point of law but also to support policy statements or provide "legal analogies" that make both rational and rhetorical appeal to readers.[248]

In a manner different from other sources of formal defects in citations, such as human error, AI-assisted legal drafting has the potential to eliminate the drafter's intentional use of citations towards any of the above ends, be it persuasion, substantiating reasoning, or serving merely as legitimation. Yet as a practical matter, citations that effectively convey a reference to a reader will generally be treated as meaningful citations regardless of their provenance. The preceding discussion then suggests

---

[242] Margie Alsbrook, *Untangling Unreliable Citations*, 37 GEO. J. LEGAL ETHICS 415 (2024).

[243] *See, e.g.*, Frank B. Cross, James F. Johnson II, Timothy R. Johnson & Paul J. Wahlbeck, *Citations in the U.S. Supreme Court: An Empirical Study of Their Use and Significance*, U. ILL. L. REV. 489, 490 (2010); John Henry Merryman, *supra* note __.

[244] *See also* Deborah L. Heller, *To Cite or Not To Cite: Is That Still a Question?*, 112 LAW LIBR. J. 393 (2020) (discussing history of unpublished opinions that are not to be cited).

[245] Cross et al., *supra* note __, at 493; Brian N. Larson, *Precedent As Rational Persuasion*, 25 LEGAL WRITING 135 (2021).

[246] *Id.* at 491; David J. Walsh, *On the Meaning and Pattern of Legal Citations: Evidence from State Local Wrongful Discharge Precedent Cases*, 31 LAW & SOC'Y REV. 337 (1997).

[247] *See* Elizabeth C. Tippett et al., *Does lawyering matter? Predicting judicial decisions from legal briefs, and what that means for access to justice*, 100 TEX. L. REV. 1157 (2022). Larson, *supra* note __, at 135, 141.

[248] Larson, *supra* note __, at 141, 143.

both that attorneys and judges using citations from AI tools may be delegating a more substantial undertaking than they may have thought, and that significant *effects* may result regardless of the drafter's (lack of) intent. Taking into account past scholarship on the significance of citations in legal proceedings, to uncritically relinquish the citations in one's drafting to an AI tool is arguably to delegate an essential element of lawyering or of recording judicial reasoning — even supposing the AI tool produces only valid citations that support correct statements of law — which may have significant but unknown implications for the persuasiveness of legal arguments (for litigant drafters), and for precedent building and for the legitimacy of the judicial system (for judicial drafters).

E. Who Bears the Burden? Who Should Bear the Burden?

Beyond choices of software design, questions about how to police citation validity bear on choices of institutional design. What is the best way to deter AI misuse in litigation, and who is best positioned to take effective action? The current state of affairs produces a great deal of busywork for courts and opposing counsel.[249] Almost certainly, a better arrangement would entail more effective advance screening before filings reach chambers. That screening should of course be undertaken by filers themselves, but it could also be undertaken by institutional actors like clerks of court, who already screen filings for compliance with certain formal requirements and bounce defective submissions back for correction.[250] Because third parties can only guess at the intent underlying an erroneous citation, optimal interventions will incentivize and facilitate preemptive error-correction by the party submitting the citation. Automated tools for detecting citation errors may be able to reallocate more corrective work to filers instead of dumping the responsibility on reviewing courts and opposing counsel.

Furthermore, as discussed above, historical approaches to legal citation have varied in their stringency. England's strict pleading standards of yore, for instance, penalized formal imperfections far more harshly than

[249] *See, e.g.*, United States v. Debaere, No. CR 4:19-CR-00009, 2026 WL 185125, at *3 (W.D. Va. Jan. 23, 2026) ("It is not only misleading to the court at the outset, but ultimately a waste of the court's time and resources, trying to sort through the many fictitious citations contained within his filing to find a kernel of support for his arguments.")

[250] *See, e.g.*, "I received a notice that I have a 'deficient' pleading, what does that mean?," U.S. BANKR. CT., D. OR., https://www.orb.uscourts.gov/faq/i-received-notice-i-have-deficient-pleading-what-does-mean (last visited July 28, 2026).

contemporary American courts generally do.[251] As tools for detecting erroneous citations improve and the institutional burdens of erroneous citations increase, it may be appropriate for modern courts to recalibrate to stricter standards. Practically speaking, it is much easier for courts to insist upon higher standards of formal perfection from litigants—and to enforce those standards with legal and professional penalties—than it is for courts to conduct forensic analyses to determine whether particular errors resulted from AI use. Such a change may help preserve institutional resources. However, more stringent citation policing would surely produce inequitable outcomes unless accompanied by improved access to legal sources and tools along the lines we have described.[252]

Conclusion

A steady bombardment of hallucinations in briefs and memoranda has motivated courts and scholars to propose new ways to police "fake" citations. These initiatives are amply justified. But the takeaway of our large-scale, empirical survey of legal filings is that identifying only true hallucinations—the "real" "fake" citations, so to speak, and not garden-variety human mistakes—is easier said than done.

Prior studies of legal hallucinations have relied upon simplifying assumptions that permit the measurement of AI models' propensity to misstate legal propositions. These findings are important and illuminating, but they deemphasize antecedent challenges that any on-the-ground mitigation of hallucinated case citations will have to address: identifying when a citation is even being made in the first place and identifying the material being cited. Our work wrangling millions of real-world citations suggests that any attempt to police hallucinations specifically will face serious impediments because, we argue, it is probably impossible to identify a hallucination from examining the suspect text alone.

However, the answer is neither to despair nor to repudiate technological assistance. We identify straightforward interventions that would almost certainly reduce the prevalence of hallucinated case citations in legal filings. Many predate generative AI: making caselaw free to access and straightforward to cite in a non-proprietary format would disincentivize under-resourced attorneys to turn to fallible chatbots and empower those attorneys to verify the representations they make to courts. Other interventions, meanwhile, may involve harnessing technology

[251] *Compare* note 71, *supra*, *with* notes 172–176, *supra*.
[252] For a discussion of potential improvements, see Parts III.A–B, *supra*.

productively: while automated citation-validation tools will probably never be able to perfectly determine whether erroneous citations are hallucinations, that limitation may not matter, because attorneys could use these tools to check their own work and correct it before it reaches third parties who have to guess about the intent behind a citation.

Hallucinated citations are indeed an institutional crisis that demands a concerted response. Our results show that any response is unlikely to be perfect. But they also show that perfection is not the standard: for all the dates in our sample, pre- and post-generative AI, legal citation has been a messy and imprecise affair, but courts found ways to make sense of the materials before them and to deter behaviors that rendered those materials incomprehensible. If it is ultimately impossible to inspect an erroneous citation and determine whether it resulted from AI use or some other human misadventure, then perhaps the best interventions will be the straightforward access-to-justice measures that help lawyers and litigants regardless of whether they use AI to write their briefs or not.