Responsible AI use for courts

Minimizing and managing hallucinations and ensuring veracity

Thomson Reuters Institute

Executive summary

Imagine a world in which AI-powered tools are not just augmenting but revolutionizing the way legal and court professionals work. This reality is unfolding rapidly, with technologies like large language models (LLM) and agentic AI driving significant positive changes in our nation's court system around areas of research, drafting, review, and case management. Indeed, this emphatic leap in technological prowess represented by AI is certain to be a huge benefit to the nation's courts, which largely remain in desperate need of a tech upgrade and slow to adopt new tech that becomes available. AI, and the promise it holds for improved efficiency and the ability to do more with less, may well be the boost that courts and their professional staff need.

In fact, as the legal profession as a whole continues to evolve, one thing is certain: AI is no longer a peripheral tool, but a central aspect of legal practice that demands attention and understanding. AI clearly has a seat at the legal table, and the conversation now must turn to using it responsibly as its adoption and application grows.

Courts and legal professionals need to transition from conducting informal exploration of AI to implementing a structured, documented program that defines when and how AI-based tools are used.

This means that courts and legal professionals need to transition from conducting informal exploration of AI to implementing a structured, documented program that defines when and how AI-based tools are used. A fundamental component of that program is understanding *hallucinations*, the phenomenon in which AI-based tools produce information that is presented with apparent confidence as accurate or factual, but is in fact incorrect, incoherent, or entirely fictitious. Hallucinations occur for various reasons, including trying to please users. Indeed, research from Princeton University indicates that the tendency of large language models to produce responses aimed at satisfying user expectations can significantly compromise factual accuracy.[1]

Understanding hallucinations requires considering the source and controls behind the information being generated. For example, general purpose consumer-grade AI tools are much more likely to produce hallucinations and/or incorrect answers than are professional-grade vertical AI solutions. Choosing a professionally managed GenAI tool with vertical depth that is designed to be used by the legal profession can help reduce hallucination risk in court settings.

As the adoption of these systems increases, their inability to confirm the veracity of their outputs may become more pronounced, which makes addressing this aspect thoughtfully even more important to help ensure quality outcomes for clients, courts, and the integrity of the judicial system.

This effort is critical, because despite AI's promise to revolutionize the legal landscape by offering unprecedented efficiency in research, drafting, and case management, AI hallucinations threaten to undermine trust and the integrity of judicial processes. While these instances highlight important considerations for implementation, they are simply a challenge that legal professionals are uniquely equipped to address through established practices of evidence review and corroboration.

The task at hand then becomes building appropriate guardrails that harness the speed and analytical power of AI while reducing the likelihood and impact of such hallucinatory mistakes. This report will focus squarely on that task: explaining what hallucinations are, why they occur, and how to design practical safeguards to ensure a more just and efficient court system.

1 *Machine Bullshit: Characterizing the Emergent Disregard for Truth in Large Language Models*; Princeton University (arXiv:2507.07484v1 [cs.CL] 10 Jul 2025); Kaiqu Liang (Princeton), Haimin Hu (Princeton), Xuandong Zhao (UCBerkeley), DawnSong (UCBerkeley), Thomas L. Griffiths (Princeton), and Jaime Fernández Fisac (Princeton).

This report demonstrates that these challenges are far from insurmountable. Our investigation, drawing insights from legal and technology experts, reveals that the tendency of AI to *guess the next word* rather than ascertain truth necessitates a proactive, human-centric approach. The solution lies not in avoiding AI, but in implementing human oversight and strategic technological checkpoints throughout the AI lifecycle.

One pragmatic path forward involves transforming hallucinations from liability into a manageable aspect of responsible AI development. This can be done by emphasizing that proactive validation and integrated verification mechanisms allow AI to deliver on its promise of unparalleled efficiency and insight, while upholding the accuracy and reliability essential to the legal system.

The core principle is clear: AI must serve as an assistant, not an arbiter, allowing human professionals to remain the ultimate arbiters of accuracy, responsible use, and decision-making when AI is involved. Regardless of the tools society creates, ultimately the integrity of the justice system will remain in the hands of the practitioners, the true stewards of the justice system, and not in the tools they use.

Methodology

This report draws upon insights from 17 interviews conducted in November and December 2025 with subject matter experts across the United States and Canada. The interview cohort included nine judges and judicial officers, two state bar attorneys, three legal subject matter experts, and three technology subject matter experts, each selected for their relevant expertise regarding AI in legal contexts.

Interviewees represented a range of technical proficiency levels and brought diverse professional perspectives to the discussion. All participants contributed in their personal capacity, and their views do not represent the official positions of their respective courts, organizations, or employers.

You can see a full list of interviewees in Appendix 1

Definitions

- **AI (artificial intelligence)** — AI refers to computer systems designed to emulate human cognitive processes through algorithms and computational methods. These systems analyze data, identify patterns, and apply learned rules to execute tasks traditionally requiring human judgment and reasoning. Core AI capabilities include processing natural language, solving complex problems, making decisions based on available information, and refining performance through exposure to new data sets.
- **Bias** — This is sometimes referred to as algorithmic bias, and it most often refers to the occurrence of unwanted or less than optimal results due to human or other biases that skew the original training data or AI algorithm. Biased AI systems may perpetuate or amplify existing societal prejudices related to race, gender, socioeconomic status, or other protected characteristics, potentially influencing decisions such as sentencing, bail, case predictions, or resource allocation. These biases can originate from historical data that reflect past discrimination, from the selection of features used in algorithms, or from design choices made by developers.[2]
- **Generative AI (GenAI)** — Unlike discriminative AI, which categorizes or makes predictions, GenAI is a subset of AI in which AI systems (often using Large Language Models) generate new content, such as text or code, based on patterns learned from vast training data. In law, GenAI platforms, such as ChatGPT-4, can draft contracts, summarize case law, and assist with legal research, significantly speeding up the work required with writing and analysis.
- **Hallucination** — This refers to the phenomenon in which GenAI-based tools produce information that is presented with apparent confidence as accurate or factual, but is in fact incorrect, incoherent, or entirely fictitious. Such outputs may include fabricated facts, citations, legal or technical authorities, code, or historical events.
- **Human in the Loop** — This refers to a collaborative approach in which humans actively work with AI systems, integrating human intelligence, feedback, and oversight into the system's learning process. This continuous loop of learning and refinement improves accuracy, reliability, and ethical decision-making.
- **Large Language Models (LLMs)** — LLMs are a type of AI system trained on vast amounts of text data that can generate human language. LLMs work by predicting the most likely next word in a sequence, enabling them to perform tasks such as answering questions, drafting documents, translating text, and summarizing content. These systems are built on complex neural network architectures; and they learn language patterns from their training data rather than accessing real-time information or understanding content in the way humans do.

2 *A Primer on the Different Meanings of "Bias" for Legal Practice*, Tara S. Emory, Esq. and Maura R. Grossman, J.D., Ph.D.; To appear in vol. 109, no. 1 of Judicature (2026).

The AI challenge before the courts

In its vital role as the third branch of government, the judiciary upholds the cornerstone principles of justice, demanding unwavering competence, transparency, accountability, and diligent oversight. The accelerating integration of AI-driven tools into legal practice now brings these foundational duties into sharp focus.

While clearly, the advent of more advanced AI-tools into the legal profession and into the courtroom specifically represents a tremendous jump in the enhanced capabilities of court professionals, it is also important not to lose sight of potential challenges.

For example, judges must be aware of AI inaccuracies and hallucinations as technology evolves because such errors pose a direct and profound threat to the integrity of judicial decision-making. This situation compels courts and their professionals to come up with a robust and responsible approach to AI's deployment within the legal system.

To address this challenge, however, judges and legal professionals must first understand it. For example, the core challenge of generative AI (GenAI) lies in its fundamental design. GenAI tools are built to predict expected language, a characteristic that inherently renders them susceptible to inaccuracies, explains Mark Francis, a leading cybersecurity, data privacy, and intellectual property (IP) attorney at Holland & Knight, adding that this predictive tendency, often shaped by user expectations, underscores the necessity for vigilant and responsible engagement to mitigate unintended biases and uphold the integrity of legal outcomes.

Further, notes Thomson Reuters legal technology expert Pablo Arredondo, LLMs are "anchored in a guess-the-next-word. AI researchers argue the level of understanding that bestows, but most agree AI doesn't have the same concept of truth that a human does." Arredondo points out that some researchers dislike the term hallucination, as "it's the exact same process that leads to hallucinations as leads to good answers. It's just whether or not a human determines the AI is correct."

This explanation reveals that AI's capacity for error is not a bug, but rather an extension of an intrinsic aspect of its function, creating new content. This further underscores the urgent need for comprehensive safeguards and human oversight.

The access to justice imperative

The legal system's inherent complexity creates a significant barrier to justice. Lawyers and judges require years of specialized training to navigate procedural requirements competently — and even with that expertise, they must exercise constant vigilance. For individuals without formal legal education, these barriers often prove insurmountable. The result is a persistent access to justice crisis that disproportionately affects those who cannot afford legal representation.

"I believe we have a serious access to justice problem in this country, and if we hold these [AI] tools to such a high standard that we never leverage them... we're doing ourselves and the thousands of people that go without lawyers a huge disservice," says Judge Maritza Braswell, a U.S. Magistrate Judge in the District of Colorado. This perspective recognizes that the risk of AI hallucinations, while real, must be weighed against the existing harm of systematic exclusion from legal processes for literally millions of citizens.

Crucially, AI tools present a potential pathway to democratize fuller legal access. However, Judge Pamela Gates, of the Maricopa County Superior Court in Arizona, says that may often not be enough.

"AI is giving individuals more access. I don't know that it's giving them an increased level of justice." This distinction is critical — technology can reduce procedural barriers and make legal information more comprehensible, but access alone does not guarantee equitable outcomes in justice system.

Further, not all AI tools are equally suited to legal applications. Dr. Maura R. Grossman, a research professor in the School of Computer Science at the University of Waterloo, warns that general-purpose systems pose unique risks. "These [public AI platforms like ChatGPT] are not proper tools for legal research… if it's something they've heard of a lot like *Roe v. Wade*, chances are you're going to get a right answer. But if you just ask for a random Tennessee case about nuisance or something like that, it will probably make it up." This observation underscores the critical need for appropriate verification mechanisms to be built into specialized tools, not simply directing unrepresented litigants to freely available general AI models.

The challenge, then, is to bridge the gap between access and justice. Through deliberate development of verification safeguards, comprehensive user education, and appropriate oversight mechanisms, AI tools can be deployed in ways that transform increased system access into meaningful justice. Again, the goal is not simply to automate legal processes, but to ensure that broader accessibility produces substantively fairer results for those historically excluded from effective legal representation.

The impact of AI hallucinations on legal practice

Again, while AI and more specifically GenAI represent a big leap forward in the technological abilities of court professionals and judges, understanding the nature of hallucinations in GenAI systems is essential for their responsible deployment in legal settings. These errors arise because AI systems generate responses based on probabilities derived from statistical patterns identified in their training data, rather than through any genuine understanding of facts, inherent capacity to verify accuracy, or *reasoning* as humans understand that term. Indeed, LLMs operate probabilistically, predicting the most likely next word in a sequence, which means that AI systems have an inherent non-determinism that makes them fallible, explains Thomson Reuters Chief Technology Officer Joel Hron.

LLMs operate probabilistically, predicting the most likely next word in a sequence, which means that AI systems have an inherent non-determinism that makes them fallible.

This probabilistic design is, by nature, creative, says Dr. Grossman, observing that for "hallucinations, the creativity [of GenAI tools] is a feature, not a bug." These systems are designed to generate novel content, which explains why they can produce outputs that appear unique or unexpected. The AI does not deliberately fabricate information; rather, it applies complex algorithms to its training data to construct responses that best align with the user's prompt. Over time and with use, these tools may further refine their outputs to match the style, expectations, or perceived preferences of individual users.

Critically, what we call *hallucinations* reflect both the system's statistically driven processes and the perception of the human user. In fact, our tendency as humans is to interpret AI-generated content as authoritative, even when it lacks grounding in verifiable fact. Recognizing this dual dimension is fundamental to integrating AI responsibly into judicial practice.

In legal practice, hallucinations can manifest in several concrete ways, including: citing non-existent cases (or, equally troubling, citing actual cases for erroneous propositions), research papers, and other authoritative documents; inventing statistics or detailed descriptions of real-world events that never occurred; presenting internally inconsistent or contradictory information within a single response; and confidently asserting legal or factual conclusions that are, on closer examination, simply wrong.

AI, Rule 11, and attorney responsibility

Another aspect of how courts must meet the challenge of AI use in the courtroom involves their policing of attorneys and litigants who appear before them. For example, Federal Rule of Civil Procedure 11 (and its counterpart in state courts) establishes a critical threshold that every attorney and self-represented litigant must cross before filing any document with a federal court. When an attorney or litigant signs a pleading, motion, or other paper, that signature carries weight. Indeed, that signature represents a personal certification to the court, after reasonable inquiry under the circumstances, that the submission meets specific standards of integrity.

In the context of AI-generated content, Rule 11 takes on heightened significance. The certification is not merely procedural, rather it represents an attorney's or litigant's personal vouching for the accuracy and legal soundness of a submission. When AI tools generate case citations, legal analysis, or factual assertions, the ultimate responsibility for verifying this information remains exclusively with the signatory. The risk of AI hallucinations creates a direct collision course with Rule 11's certification mandate.

Although Rule 11 is a federal provision, every jurisdiction imposes a comparable duty on attorneys to investigate the factual and legal basis of their filings and to bring only meritorious claims and defenses.

Rules on technical competence

In addition to the Federal Rules of Civil Procedure, the ABA Model Rules impose baseline duties of competence that extend to the use of technology, including AI. Model Rule 1.1, adopted in some form by more than 40 US states, requires lawyers to possess the knowledge, skill, thoroughness, and preparation reasonably necessary for proper representation — and this obligation is increasingly understood to encompass a basic understanding of relevant technological tools and their attendant risks.

Likewise, Model Code of Judicial Conduct Rule 2.5 requires judges to perform their duties competently and diligently. Taken together, these provisions indicate that both attorneys and judges must develop and maintain a foundational level of technical competence with respect to AI in situations in which it materially affects litigation, adjudication, or the administration of justice.

Most of the legal experts we interviewed concluded that sweeping new rules are not necessary, because existing procedural and ethical frameworks already govern the use of AI.

Against this backdrop, most of the legal experts we interviewed concluded that sweeping new rules are not necessary, because existing procedural and ethical frameworks already govern the use of AI.[3] Lawyers and judges are required to maintain technological competence, and in contemporary practice that obligation encompasses a working understanding of how AI-based tools function. This does not require expertise in coding or software design, but it does require the ability to scrutinize AI-generated outputs and assess when (if ever) an output is sufficiently reliable to inform legal analysis, filings, or judicial decision-making.

Some states have issued further guidance. The California State Bar, for example, has articulated a particularly clear position on professional responsibilities associated with AI use. In its guidance on internal AI deployment, the California Bar emphasized that AI must be treated as an "assistive tool rather than an authoritative source." All AI-generated content is to be checked against primary materials

3 *Is disclosure and certification of the use of generative AI really necessary?* Maura R. Grossman, Paul W. Grimm & Daniel G. Brown; Judicature (Vol.107, No. 2) October 2023.

and subjected to independent human review. This framework prioritizes verification, human oversight, and responsible use as core risk management principles.[4]

This same philosophy is reflected in the Practical Guidance on AI issued by the California State Bar's Committee on Professional Responsibility and Conduct (COPRAC), which underscored that human legal judgment must always remain paramount. Under COPRAC's approach, attorneys may not rely on AI outputs without confirming their accuracy through established, reliable sources.

The TR Institute's View:

Providing access *and* justice

When it comes to ensuring both access to the legal system *and* a just outcome for unrepresented or under-represented litigants, the choice of AI platform significantly impacts reliability and thus, justice. General-purpose, open-source AI tools (such as publicly available chatbots like Chat-GPT) pose substantially higher hallucination risk compared to professional-grade, legal-specific platforms. Consumer-facing tools are optimized for general conversation rather than the precision and citation accuracy that legal proceeding and court filings demand.

Courts need to ensure the adoption of professional legal AI platforms that can incorporate critical safeguards that are absent from open-source alternatives. These professional legal AI platforms should include curated legal databases with verified case law and statutes; citation verification features that can link content to authoritative sources; jurisdiction-specific training; audit trails that document the basis for AI-generated analysis; and regular updates reflecting current law.

Even as AI-driven tools bring tremendous benefits and efficiencies to the working of the nation's courts, judges and professional staff have to make access to justice work on both counts, providing professional legal AI platforms that can offer access to under-represented litigants while also providing the AI-driven prowess to allow them to obtain a just outcome.

4 The State Bar of California; *Generative AI Governance Framework and Step-by-Step User Guide* (November 20, 2025); available at: https://shorturl.at/60MGw.

Hallucinations in the wild

Not surprisingly, publicly reported examples of AI hallucinations in legal proceedings have occurred primarily, but not exclusively, in the United States, where court records and disciplinary matters are relatively accessible. Incidents have involved self-represented litigants, paid attorneys, and judges. While the first widely publicized cases appeared in 2023, such incidents are becoming more frequent and likely reflect the rapid and expanding use of various AI tools in legal practice.[5]

Three notable cases carry heavy lessons about the potential impact of hallucinations within legal proceedings.

Mata v. Avianca: The emergence of AI-related sanctions

In June 2023, one of the first US cases sanctioning a lawyer for citing fictitious authorities, *Mata v. Avianca, Inc.*[6] became a highly publicized inflection point in the conversation about AI use in litigation matters. In that case, counsel not only relied on hallucinated case law generated by an AI tool but repeatedly vouched for its accuracy and even submitted fully fabricated excerpts purporting to quote non-existent decisions. Under review, one lawyer claimed to not know that the AI-based tool could fabricate whole cases; and a second lawyer did not properly verify the cases cited or provided to the court.

The court sanctioned both the individual attorneys and their law firm not merely to remedy the immediate misconduct, but to send a clear warning that unverified AI output cannot be passed off as legal research. As a result, the lawyers involved, and their firm, have become closely associated with the risks of filing AI-hallucinated material in court.

The impact of *Mata* extends well beyond the parties themselves. It has become a cautionary tale for the broader bar, signaling that although AI tools may handle much of the *heavy lifting* in research and drafting, they cannot substitute for a lawyer's professional judgment. The core tasks of legal reasoning, issue spotting, and applying the law to fact remain squarely with the attorney. The central lesson from *Mata* is that accountability resides with the *humans in the loop* — those humans actively working with AI systems by integrating human intelligence, feedback, and oversight into the process. To this end, lawyers must verify, contextualize, and be prepared to stand behind any AI-assisted work product they file in litigation.

Thomas v. Pangburn: Balancing compassion and accountability

When dealing with self-represented litigants who use AI, courts must balance accountability with an appreciation for how daunting it can be to navigate the legal system without counsel and without sacrificing the integrity of the system, explains Justice Tanya R. Kennedy, an Associate Justice of the Appellate Division, First Judicial Department of New York. This became clearer than ever when evaluating the judge's response in another case, *Thomas v. Pangburn*.[7]

The court's posture illustrated that self-represented litigants may rely heavily on online tools and secondary sources. They may be unclear, or even oblivious, about the true origins and reliability of the citations and materials they submit. In *Thomas*, the litigant submitted a pleading with several hallucinated authorities and, upon an Order to Show Cause, failed to explain how such errors occurred. In such circumstances, judges are tasked with treating self-represented litigants with dignity, explaining

5 *AI Hallucination Cases* database; Damien Charlotin; available at: https://www.damiencharlotin.com/hallucinations/?q=&sort_by=date&period_idx=0.

6 *Mata v. Avianca, Inc.*, 1:22-cv-01461 (PKC) (S.D.N.Y. 2023).

7 *Thomas v. Pangburn*, CV423-046, 2023 WL 9425765 (S.D. Ga. Oct. 6, 2023).

why hallucinated or unreliable AI-generated content is problematic, and, where appropriate, offering clarification or an opportunity to correct the record — all while ensuring that judicial resources, particularly time and focused attention, are not squandered on frivolous or misleading filings.

In *Thomas*, sanctions were ultimately deemed appropriate because, after full consideration, the court concluded that the self-represented litigant had acted in bad faith by "delaying or disrupting the litigation." That decision underscored that compassion does not eliminate the need to protect the integrity and efficient functioning of the court.

Shahid v. Esaam: A watershed AI moment

Prior to 2025, the legal community derived measured reassurance from a critical distinction — that the judiciary itself had not made the mistake of relying on a hallucination. Although no one considered judges infallible, there remained an underlying assumption that judicial discipline, along with the assistance and training provided served as inherent protections against uncritical acceptance of AI-generated material.

The decision in *Shahid v. Esaam*,[8] disrupted that assumption. In this case, a judge incorporated AI-hallucinated citations into a judicial order after trusting content drafted by one party's attorney without independent verification. This incident fundamentally altered what could be presumed reliable in legal proceedings — extending doubt even to the contents of court orders themselves.

The paradigm shift this case represents: The emerging norm has moved from "trust but verify" to "do not trust until verified.

Dr. Grossman describes the paradigm shift this case represents: The emerging norm has moved from "trust but verify" to "do not trust until verified." *Shahid* demonstrated what many judges had privately feared — that no group, regardless of experience or institutional position, is immune to the persuasive presentation of AI-generated hallucinations.

The cost of creation: Sanctions for AI hallucinations

When AI-generated hallucinations are discovered in the context of the courts, sanctions generally follow. The nature and severity of these sanctions vary considerably, with courts calibrating their response based on the circumstances and the party involved, similar to the response illustrated in *Thomas*.

In practice, this may mean that for a first offense, courts issue a warning, cautioning litigants about their obligation to verify AI-assisted filings and, where appropriate, offering an opportunity to correct the record before considering more serious steps such as striking sections that rely on hallucinated authorities. This educational and corrective approach recognizes both the ubiquity of AI tools and the reality that self-represented litigants may not fully grasp those tools' limitations, while still underscoring that misuse burdens the system and erodes trust.

Legal practitioners, however, are held to a different standard. "Whether you are a judge [or] an attorney, credibility is everything, particularly when you come before the court," says Justice Kennedy, adding that genuine harm can be done by lawyers who invoke AI irresponsibly, either by submitting hallucinated content or by making unfounded insinuations that opposing counsel has relied on GenAI without any factual basis. In Justice Kennedy's

"Whether you are a judge [or] an attorney, credibility is everything, particularly when you come before the court."

– Justice Tanya R. Kennedy
Associate Justice of the Appellate Division,
First Judicial Department of New York

8 *Shahid v. Esaam*, 918 S.E.2d 198 (Ga. Ct. App. 2025).

view, such conduct is “not only concerning for the court, but also for even the attorney’s client” and raises serious questions about civility and professional responsibility.

Reflecting this heightened duty, courts are increasingly requiring attorneys to specify how AI was used in preparing submissions and are imposing more substantial sanctions when misuse is found, including mandatory Continuing Legal Education (CLE) courses on ethics and technology and, in some cases, referring the matter to grievance committees. This elevated accountability for practitioners mirrors their professional obligations to verify work product, uphold the integrity of the system, and protect clients’ interests.

While the imposition of sanctions underscores the seriousness of AI misuse, these measures also serve an educational function, establishing clear standards for responsible AI integration and fostering greater awareness of both the technology’s potential and its limitations.

Despite advances in human oversight protocols and technological safeguards, AI hallucinations will remain an ongoing risk in legal practice. Cases involving hallucinations typically reveal attorneys who would likely have exhibited lapses in due diligence regardless of the tools employed. Yet these incidents serve as critical teaching moments that justify both corrective action and deterrent measures.

“Three years down the line you will still have self-represented litigants asking any kind of free large language models on the Internet to write a brief — and there will be hallucinations, and you cannot prevent that,” says Damien Charlotin, a Research Fellow in law, data science, and AI at HEC Paris. Indeed, he adds, the accessibility of AI tools means that hallucinated content will continue entering court filings from all participants in the process.

Given this reality, enforcement mechanisms become essential. Charlotin emphasizes that existing professional obligations already provide the framework: “Lawyers, attorneys, [and] self-represented litigants should check their references because they’ve got already a duty to do that.” Already many major legal research providers offer tools that will check a brief, court order, or opinion for accurate case citations and quotations.

When these duties are breached, however, courts must respond decisively. Judge John Blanchard, of the Maricopa County Superior Court in Arizona, describes the balanced approach required, saying: “Calling hallucinations out is important, either in a Show Cause Hearing or a firmly worded minute entry; but education is key, and we’re doing everything we can over here to develop that as well.”

The TR Institute’s View:

Treating verification as part of workflow

While the novelty of AI and the potential of its beneficial impact may change the scale and detectability of errors, it does not change the duty to verify. Courts should treat verification as a workflow requirement, not an afterthought, and hold those that come before them — either lawyers or self-represented litigants — to appropriate standards. Of course, these standards should make a distinction in the level or type of sanctions imposed on lawyers who have access to more reliable legal tools and self-represented litigants who may not.

Courts also should not shy away from imposing sanctions, such as monetary penalties, disgorgement of fees, disciplinary hearings, and mandatory training, in order to hold individual practitioners accountable for failures in due diligence. These corrective and at times punitive actions also establish clear precedents that educate the broader legal community about verification requirements in AI-augmented legal practice.

Implementing effective safeguards

Despite the potential for efficiency gains and impactful workflow improvements — or, perhaps because of them — the stakes around AI usage in court remain high. That means the current environment necessitates a proactive and robust approach to safeguards within the judicial system. True protection against the potential pitfalls of AI, particularly hallucinations, demands that human actors are strategically positioned and properly equipped.

Those entrusted with reviewing AI-generated content must possess both the appropriate verification tools and comprehensive training. This ensures they can effectively exercise meaningful oversight and definitively confirm the accuracy of all information presented to and disseminated by the court. The goal is not merely to have a human glance at AI output, but for that human to be capable of rigorous scrutiny and independent validation — this is what is meant by a *human in the loop*.

As Judge Erica R. Yew of the Superior Court of Santa Clara County in California notes, some of the apprehension surrounding AI stems from a misconception that the technology introduces entirely new problems. “Some of the angst that people are having is they think some of these issues are new, but these are the same issues courts have dealt with for so long,” explains Judge Yew. “It’s just that they’re more pervasive and sometimes harder to detect.” While AI may make inaccuracies more widespread or subtle, the core challenge of ensuring truthful information is not novel to the legal system.

Judge Yew further emphasizes that existing mechanisms remain vital: “The Rules of Evidence work to help judges and lawyers ferret out things that shouldn’t be admissible. Careful research and the judge or their staff checking the citations that lawyers provide has always been important. And a lot of the safeguards that our legal system has in place still work and will continue to work going into the long term.”

Her perspective underscores how while AI presents unique dimensions to old problems, the fundamental human-centric safeguards embedded within our legal processes — such as adherence to rules of evidence and diligent verification by judges and their staff — are crucial and continue to provide a strong foundation for maintaining accuracy and trust.

The persistent need for human oversight

Judge Braswell agrees. “Just because the concept of AI hallucinations is new, it doesn’t mean the issues underlying it are new,” she says. “That same diligence that was required before the new AI tech is still the same diligence required now.” This observation reframes the issue significantly. Indeed, the fundamental error in cases involving AI hallucinations is not the technology’s malfunction itself, but rather the failure to apply adequate human oversight. When viewed through this lens, improper reliance on AI tools parallels the consequences of depending on inadequately supervised support staff or junior associates. The ultimate responsibility for accuracy remains with the litigant, attorney of record, or the judge that signs off.

The California Bar emphasized this principle explicitly, noting that no technology employed in legal practice — whether AI-powered research tools, traditional search engines, or document automation systems — achieves *complete accuracy*. The critical factor is not the tool’s perfection, but rather the practitioner’s approach: Attorneys

“Just because the concept of AI hallucinations is new, it doesn’t mean the issues underlying it are new. That same diligence that was required before the new AI tech is still the same diligence required now.”

– Judge Maritza Braswell,
U.S. Magistrate in the District of Colorado

must comprehend the inherent limitations of their tools, independently verify all outputs, and apply their professional judgment before relying on any technologically generated content.

Technology as a verification partner

Fortunately, advanced technology itself can offer a path to solution. In fact, the very AI technology that is bringing heightened efficiency and enhanced workflow benefits can and should play an active role in helping detect and mitigate hallucinations within AI systems. Human oversight remains indispensable, of course, but expecting manual verification of every AI-generated citation, factual assertion, or legal proposition is unrealistic at scale. To be effective, human decision-makers must be equipped with targeted tools that can identify likely errors quickly and explain why something appears problematic, while integrating naturally into existing legal workflows.

Today, a growing ecosystem of AI-driven tools is emerging to meet this need. At one end are verification tools that automatically check citations against authoritative databases, flagging nonexistent cases, misquoted passages, or mischaracterized holdings. Others focus on fact-checking, comparing AI outputs against trusted sources such as court records, statutes, and reputable secondary materials. More specialized systems are designed to operate as *AI auditors*, running in parallel with GenAI tools to detect internal inconsistencies, missed reasoning steps or claims that deviate from known legal doctrine. Document management platforms are also beginning to embed AI that tracks the provenance of text, making it easier to see what content was drafted by humans, what was generated by AI, and what has been independently validated.

The practical challenge for courts then is to select and combine tools in a way that aligns with their institutional role, risk tolerance, and resource constraints. A high-volume court handling self-represented litigants may prioritize simple, automated checks that flag obviously spurious citations, while an appellate court may rely on more advanced analytical tools to scrutinize complex doctrinal arguments. In each context, the goal is to create a layered system in which human judgment is supported by technology that makes hallucinations more visible, more explainable, and ultimately less likely to influence legal outcomes.

The TR Institute's View:

Remaining under human supervision

Clearly, AI technologies offer substantial benefits across the legal ecosystem. When deployed responsibly, these tools deliver measurable efficiency gains, freeing legal professionals to focus more time on substantive analysis and client service.

However, realizing these benefits requires understanding a fundamental principle: AI must function as a tool under human supervision, never as an autonomous decision-maker. Overreliance on AI output — particularly accepting it without independent verification — amplifies the inherent risks of probabilistic systems that generate responses based on pattern prediction rather than factual accuracy.

To mitigate this risk, courts and legal practitioners should prioritize professional-grade, legal-specific AI tools over general-purpose alternatives. The investment in verified legal technology platforms is prudent compared with the professional liability risks, reputational harm, and potential sanctions associated with submitting hallucinated content during legal proceedings.

One key principle: AI is an assistant, not an arbiter

Given the potential for the courts to take a technological leap forward with increasing use of AI, it is important that a foundational principle has emerged for its responsible use in legal practice: *AI should serve as an assistant, not an arbiter*. This principle emphasizes that technology must enhance, rather than replace, the critical reasoning and professional judgment that define effective legal work.

Indeed, the full benefits of such advanced technology will only come to fruition if it's paired with responsible use and carefully applied safeguards.

"If all you do is take information that generative AI produces, drop it into a brief, and sign it, you have misused the tool," warns Judge Samuel A. Thumma, Division One of the Arizona Court of Appeals, adding that AI may provide a starting point, but human analysis must begin there, not end. The consensus among judges is equally unequivocal — counsel cannot simply submit AI-generated material to a court without rigorous verification and independent analysis.

"If all you do is take information that generative AI produces, drop it into a brief, and sign it, you have misused the tool."

– Judge Samuel A. Thumma, Division One of the Arizona Court of Appeals

Judge Maryam Ahmad, an Associate Judge in the First Municipal District of Chicago, expands on AI's potential roles, describing it as "a tool with multiple uses" that can function as a brainstormer or provide insight on specific subjects. Yet she reinforces a critical caveat, noting that "no matter what information you get, it doesn't relieve you of the responsibility to check and double check." The versatility of AI as a legal tool does not diminish the practitioner's ultimate responsibility for accuracy and validity.

Beyond drafting assistance, AI can serve as what Thomson Reuters' Hron describes as "a thought partner or like a critic" that tests preconceived notions and biases. This function — helping to "open the aperture of viewpoints on a particular topic" represents one of AI's most promising applications in legal reasoning. By presenting multiple perspectives and options, AI can help practitioners arrive at more robust conclusions, provided the technology is used as a catalyst for deeper analysis rather than a substitute for it. This collaborative approach — between humans and AI — is increasingly recognized by judicial officers as a valuable dimension of AI's role in legal practice. "AI should augment us, not replace us," says Judge Braswell. "Right now, the focus is on automation and having AI take over tasks. The more productive focus is augmentation — using AI to surface better information, test assumptions, identify gaps, and support *our own* decision-making. That shift will strengthen us."

By presenting multiple perspectives and options, AI can help practitioners arrive at more robust conclusions, provided the technology is used as a catalyst for deeper analysis rather than a substitute for it.

Across diverse legal settings, from courtrooms to law firms, a consistent message prevails: While AI use can provide tremendous benefits, its purpose is to support, not to supplant, the exercise of legal judgment. The very legitimacy of judicial decisions hinges on the independent reasoning of experienced legal professionals who meticulously examine evidence, weigh precedents, and apply legal principles to specific factual circumstances.

Within this framework, AI tools are most effectively conceptualized as junior research assistants. “I treat AI as a junior research assistant, not a decision maker,” notes one federal judge. “Anything AI suggests must be confirmed in authoritative databases or the case record before it influences my reasoning or appears in an order. Nothing goes into a final decision unless I can independently tie it to the record and real, verified law.”

The TR Institute’s View:

The human-AI collaboration

Despite its ability to greatly enhance efficiency and cost-effectiveness (especially in a tech-impaired and overburdened court system) AI still fundamentally lacks the capacity to serve as judicial decision-makers or definitive arbiters of legal or factual determinations. That essential responsibility resides exclusively with human judges and legal professionals, who possess the requisite combination of substantive legal knowledge, ethical judgment, contextual reasoning, and individual accountability that no advanced technology can replicate.

However, maintaining human primacy does not require abandoning technological assistance in the verification process itself. Verification must be integrated throughout the entire AI lifecycle, not merely applied as a final check before filing. This means human authorities need to establish validation checkpoints at multiple stages, such as when formulating prompts, during initial output review, and before final submission. At each stage, legal professionals should employ professional-grade research tools to confirm AI-generated citations, cross-reference legal propositions against authoritative databases, and validate factual assertions.

This human-AI collaboration, which leverages the exercise of human judgment and is supported by appropriate verification tools and workflows, remains indispensable to preserving both the integrity of judicial proceedings and the constitutional rights of all parties appearing before the court.

Weighing the benefits and risk going forward

While the issue of AI hallucinations presents certain challenges within legal proceedings, courts and law firms should not be deterred from utilizing these advanced technologies. Indeed, the promise of tech-driven efficiency and cost-savings to legal practice and the courts, as well as the potential benefits to access to justice, means that when weighing the benefits and risks of AI use, these advantages should be considered top-of-mind. However, the awareness of hallucinations also should guide the strategies employed in its implementation.

The solution to the hallucination conundrum lies in implementing robust verification protocols that specifically remain under human oversight. This is essential at every stage of AI-assisted legal workflows. This oversight may take the form of comprehensive review throughout the research and drafting process or systematic validation of final outputs; in any case, the critical requirement is that all AI-generated legal content undergoes independent verification by qualified individuals.

While AI systems can generate legal analysis and citations with remarkable fluency and apparent authority, legal professionals must resist the tendency to accept outputs uncritically. The consequences of unchecked AI hallucinations in court underscore the necessity of maintaining rigorous professional standards. To harness AI effectively as a productivity tool while upholding the integrity of legal practice, courts and law firms should take the following actions:

- Establish mandatory review procedures for all AI-generated work products, including briefs, memoranda, and research summaries
- Train legal personnel to approach AI outputs with appropriate professional skepticism and an understanding of how hallucinations occur
- Implement fact-checking protocols that verify case citations, statutes, and legal precedents against authoritative legal databases
- Designate experienced attorneys or subject matter experts to validate complex legal arguments and specialized content
- Utilize specialized legal AI tools designed with built-in verification features and citation validation capabilities
- Document verification processes to ensure accountability, quality assurance, and compliance with professional responsibility standards
- Develop clear policies regarding AI use that align with evolving ethical guidelines from bar associations and judicial bodies

By treating all AI as a collaborative assistant rather than an autonomous agent, legal professionals can capture the vast efficiency gains these technologies offer while preserving the accuracy and reliability that the practice of law demands. The goal is not to avoid AI, of course, but to integrate it responsibly within human-supervised workflows that maintain the profession's commitment to truth, accuracy, and justice.

Ultimately, the prevention of AI hallucinations in court depends on the continued exercise of human judgment, professional expertise, and ethical responsibility at every stage of the legal process.

The TR Institute's View:

Building on the foundation we have

Undoubtedly, AI can significantly enhance the quality and efficiency of legal work, and the path forward is clear and achievable. To that end, courts and legal professionals need to establish strategic AI implementation processes that are supported by comprehensive education, robust verification workflows, and transparent accountability for AI-assisted work.

Fortunately, the legal profession doesn't need to start from scratch, because its existing ethical frameworks and professional standards around competence, due diligence, and candor to the court provide an excellent foundation for AI integration. By building on these proven standards and implementing proper verification protocols, lawyers and judges can confidently harness AI's capabilities for rapid legal research and document analysis while maintaining professional rigor, allowing them to focus more deeply on the analytical reasoning, ethical judgment, and human understanding that remain uniquely within their domain.

For further insight, please read the Frequently Asked Questions about AI hallucinations in Appendix II below.

Appendix I: List of source experts interviewed

The interviewees for this report represented a range of legal and technical proficiency levels and brought diverse professional perspectives to the discussion. All participants contributed in their personal capacity, and their views do not represent the official positions of their respective courts, organizations, or employers.

1. **Judge Maryam Ahmad** — Circuit Court of Cook County, Illinois; First Municipal Division.
2. **Pablo Arredondo** — Vice President, CoCounsel, Thomson Reuters
3. **Judge John Blanchard** — Maricopa County Superior Court in Arizona
4. **Judge Maritza Braswell** — United States Magistrate Judge in the District of Colorado
5. **Rachel Brewer** — Managing Attorney, Office of Professional Competence, The State Bar of California
6. **Damien R. Charlotin** — Researcher in law, data science, and AI at HEC Paris
7. **Mike Dahn** — Head of Westlaw Product Management, Thomson Reuters
8. **Mark Francis** — Partner, Holland & Knight
9. **Judge Pamela Gates** — Maricopa County Superior Court in Arizona
10. **Dr. Maura R. Grossman** — Research Professor in the School of Computer Science at the University of Waterloo (Canada), and Adjunct Professor at Osgoode Hall Law School of York University (Canada)
11. **Joel Hron** — Chief Technology Officer, Thomson Reuters
12. **Justice Tanya R. Kennedy** — Associate Justice of the Appellate Division, First Judicial Department of New York
13. **Catherine Ongiri** — Program Director, Office of Professional Competence, The State Bar of California
14. **Commissioner Kendra Thomas** — Commissioner with a courtroom in the Los Angeles Superior Court
15. **Judge Samuel A. Thumma** — Division One of the Arizona Court of Appeals
16. **Judge Erica Yew** — Superior Court of Santa Clara County in California

Appendix II: Demystifying AI — Frequently Asked Questions

What are the potential applications of AI tools in judicial proceedings and court operations?

When implemented responsibly, AI tools can offer judges, court staff professionals, lawyers and litigants a great leap forward in terms of efficiency and effectiveness.

Indeed, AI technologies can enhance court operations in three primary domains: *i*) administrative functions; *ii*) productivity enhancement; and *iii*) legal research and drafting support. AI-powered systems can streamline court operations by providing automated assistance to litigants, attorneys, and court personnel through intelligent chatbots that facilitate case scheduling, procedural guidance, and navigation of court processes, thereby improving access to justice and reducing administrative burden. Additionally, AI tools can serve as sophisticated research and drafting assistants, enabling legal professionals to efficiently analyze case law, statutes, and legal precedents and to write more effectively. They also may function as a collaborative thought partner during case preparation, brief drafting, and strategic legal analysis, augmenting — rather than replacing — the professional judgment of trained legal practitioners.

What are the classifications of AI hallucinations and what causes them?

AI hallucinations occur when the system generates a response that is presented with apparent confidence as accurate or factual, but is in fact incorrect, incoherent, or entirely fictitious. These erroneous outputs are fundamentally rooted in human input — whether through the initial prompt provided by the user, the training data curated by developers, the design parameters established during system development, or inadequate human oversight during implementation. Ultimately, GenAI systems are making probabilistic determinations based on their training data, which can vary in quality and accuracy.

To what extent can AI algorithms and technological tools effectively prevent or detect hallucinations, and what is their reliability?

While technological solutions for detecting AI hallucinations are emerging, their effectiveness remains limited and cannot be relied upon as a comprehensive safeguard. Because these technological detection methods are subject to limitations, the legal profession cannot and should not rely exclusively on technological solutions to ensure the accuracy of AI-generated content. Detection tools may *assist* in identifying obvious errors or inconsistencies, but they do not eliminate the fundamental need for human professional judgment and independent verification. Legal practitioners and judicial officers must treat AI-assisted detection tools as supplementary aids rather than definitive validation mechanisms.

Do AI hallucinations present distinct challenges compared to human errors?

Yes. While AI sometimes makes mistakes just like a human would (for instance, failing to find a relevant law or coming to an incorrect conclusion), it often makes mistakes in ways that no human ever would, by, for example, adding a fourth element to a cause of action that only has three

elements. This can make review challenging, because if you review AI output in the same the way that you would review the output of a junior lawyer, you can easily fail to correct these kind of mistakes.

In addition, AI systems can generate output with a veneer of coherence and confidence that may reduce the natural skepticism legal professionals might otherwise apply to unfamiliar sources or questionable assertions, thereby increasing the risk of overreliance and inadequate verification.

The ultimate accountability for any error in legal work product, including hallucinated or inaccurate information generated by AI systems, rests entirely with the legal professional who submits, relies upon, or presents such content, so when reviewing AI output, remember that it can make mistakes in ways no human ever would.

When allegations of inappropriate AI-generated content arise, who bears the burden of proof?

When a party alleges that opposing counsel or another litigant has submitted inappropriate, inaccurate, or hallucinated AI-generated content, the fundamental principles of procedural fairness and professional courtesy dictate that the *identifying party* bears the initial burden of demonstrating a reasonable basis for the allegation. This is critical when allegations are made in a manner that questions professional competence or ethical conduct, as such accusations carry significant reputational consequences for legal practitioners. The identifying party must present specific evidence or articulate concrete reasons supporting a good-faith belief that the challenged content was AI-generated and contains material errors, rather than making speculative or blanket assertions designed to harass opposing counsel or delay proceedings. Courts should be vigilant in preventing the weaponization of AI-related allegations as a litigation tactic if they are absent substantial justification.

How might hallucinations in AI-generated content influence judicial determination or compromise due process?

When employed appropriately as a research or analytical tool subject to rigorous human oversight, AI-generated content should not adversely affect judicial determinations or cause due process concerns. The critical safeguard lies in maintaining the judicial officer as the ultimate decision-maker, with AI serving solely as an assistive instrument rather than a substitute for independent judicial analysis.

Potential risks to due process arise only when judicial officers abdicate their responsibility to critically evaluate AI-generated materials or rely upon such content without adequate verification. To mitigate these risks, judges and judicial staff must:

- treat AI as a supplementary tool that enhances, but does not replace legal research and reasoning;
- maintain active vigilance for potential hallucinations, including fabricated citations, erroneous legal standards, or fictitious factual assertions;
- independently verify all substantive legal authorities, factual representations, and analytical conclusions derived from AI-generated content before incorporating them into judicial work product; and
- exercise the same professional judgment and critical analysis that would be applied to any other research resource.

What methodologies can courts use to validate the accuracy of AI-generated legal citations and precedents?

Courts must recognize that absolute accuracy in AI-generated content remains an unattainable standard. Accordingly, validation methodologies must focus on implementing robust verification processes that position human reviewers as the essential safeguard against inaccuracies, fabricated citations, and erroneous legal authorities.

The primary methodology for ensuring accuracy, of course, is independently verifying all AI-generated legal citations and precedents before they are relied upon in any judicial process. The reviewer should independently confirm that cited cases exist in official sources, that quotations are accurate and properly contextualized, that legal propositions reflect the actual holdings of the cited authorities, and that precedents remain good law and have not been overruled, distinguished, or superseded. Again, many major legal research providers currently offer the types of tools that will allow users to check briefs, judicial orders, and opinions for accurate case citations and quotations.

Obviously, that is an awesome undertaking for already overburdened courts and professional staff. That's why beyond human oversight, courts should consider investing in technological solutions that incorporate built-in accuracy safeguards. This may include AI systems with integrated citation verification features that cross-reference legal databases in real time, tools that flag potentially problematic outputs for heightened scrutiny, or platforms that provide confidence scores or source links for generated content.

What educational and training initiatives should judicial bodies implement to mitigate or prevent hallucinations?

Judicial bodies should establish comprehensive educational programs ensuring that all individuals who interact with the legal system possess foundational knowledge of AI and its appropriate application in legal contexts. While technical expertise is not required, a baseline understanding of AI fundamentals is essential for responsible integration of these tools into judicial processes.

At a minimum, educational initiatives should address the basic operational principles of GenAI, including the concept of prompts and the predictive, probabilistic nature of LLMs. Participants must understand the inherent limitations of AI technology, as well as the critical distinction between AI as a research tool and AI as a decision-making authority.[9]

How can AI, particularly LLMs, be responsibly deployed to better enhance public access to justice and support self-represented litigants?

When implemented with appropriate safeguards and human oversight, AI-driven tools and LLMs offer significant potential to expand access to justice by addressing longstanding barriers that have historically impeded meaningful participation in the legal system for many citizens. These technologies can serve as force multipliers for self-represented litigants, democratizing access to legal processes that have often been prohibitively complex or resource intensive.

For self-represented litigants, these advanced technologies can serve as educational resources that can potentially demystify legal procedures and terminology, translating complex legal concepts into plain language that facilitates comprehension and informed participation. These tools can greatly assist individuals in understanding court rules, preparing basic filings, identifying relevant legal issues, and navigating procedural requirements that would otherwise require costly legal consultation.

9 For more on this, you can see the *AI Policy Consortium for Law & Courts*, a joint effort between the Thomson Reuters Institute and the National Center for State Courts, that provides a role-based learning tool kit that helps to enhance AI Literacy in the courts. The Consortium is available here: https://www.ncsc.org/resources-courts/artificial-intelligence-ai.

Thomson Reuters

Thomson Reuters is a leading provider of business information services. Our products include highly specialized information-enabled software and tools for legal, tax, accounting and compliance professionals combined with the world's most global news service – Reuters.

For more information on Thomson Reuters, visit **tr.com** and for the latest world news, **reuters.com**.

Thomson Reuters Institute

The Thomson Reuters Institute brings together people from across the legal, corporate, tax & accounting and government communities to ignite conversation and debate, make sense of the latest events and trends and provide essential guidance on the opportunities and challenges facing their world today. As the dedicated thought leadership arm of Thomson Reuters, our content spans blog commentaries, industry-leading data sets, informed analyses, interviews with industry leaders, videos, podcasts and world-class events that deliver keen insight into a dynamic business landscape.

Visit **thomsonreuters.com/institute** for more details.

Thomson Reuters Institute