



ARTIFICIAL INTELLIGENCE VS LEGAL ETHICS: 5 RISKS YOU CAN'T IGNORE

Swetaleena Sahoo *

INTRODUCTION

Generative artificial intelligence has moved swiftly from a curiosity in legal practice to a routine drafting aid, and courts have moved just as quickly to test its limits. Over the past two years, judges across the United States, the United Kingdom, and increasingly India and Singapore have confronted a recurring problem: lawyers filing briefs that cite cases which do not exist, misstate holdings, or quote judgments inaccurately. What began as a handful of isolated embarrassments has hardened into a documented pattern with real professional consequences. A public database tracking court decisions that address AI-hallucinated material recorded roughly 1,600 matters worldwide by mid-2026, the large majority in United States courts, with new entries added almost daily.¹ Independent research helps explain why the problem persists: a Stanford-affiliated study of AI legal-research tools found hallucination rates of around 17 per cent for one leading subscription product and 33 per cent for another, against roughly 43 per cent for a general-purpose chatbot asked the same legal questions figures that should trouble any lawyer inclined to treat AI output as reliable by default.² This note sets out five ethical risks that any lawyer using AI tools should keep front of mind, together with the professional-conduct principles that already govern them, whether or not a jurisdiction has adopted AI-specific rules.

FABRICATED CITATIONS AND THE NON-DELEGABLE DUTY TO VERIFY

The most publicised risk remains the AI 'hallucination ': a confident-sounding citation, quotation, or case summary that is simply invented. The pattern was set early: in *Mata v.*

*ADVOCATE.

¹ Steven Melendez, 'Why Lawyers Keep Citing Fake Cases Invented by AI' (Scientific American, 22 May 2026); 'AI Hallucination Legal Cases: A Sanctions Tracker' (GC AI, 2026).

² 'AI Hallucination Sanctions: A Lawyer's Guide for 2026' (TheLawGPT, 2026), citing Stanford RegLab research on AI legal-research tool reliability.

Avianca Inc., a New York lawyer submitted a brief citing six decisions that did not exist, having asked a general-purpose chatbot to find supporting authority; the court imposed a \$5,000 sanction on the lawyer, his co-counsel, and their firm.³ What might once have looked like an isolated lapse has since become a recurring feature of litigation practice worldwide.

Courts have been unambiguous that the duty to verify every citation in a filing belongs to the signing lawyer and cannot be delegated to software, a paralegal, or a junior associate, however the error occurred. The financial and professional stakes have escalated sharply since *Mata*. In an Oregon dispute, two lawyers filed briefs containing fifteen fabricated cases and eight invented quotations, resulting in combined sanctions of roughly \$110,000, reportedly the largest aggregate penalty of its kind to date.⁴ A federal court of appeals has imposed \$15,000 fines per attorney for comparable conduct.⁵ Seniority offers no shield: in one Singapore matter, the High Court imposed personal costs orders on both the junior lawyer who generated the hallucinated content and the partner who signed the filing without checking it, expressly rejecting the partner's workload as a defence.⁶

Disciplinary consequences have followed the same trajectory. A Colorado lawyer who blamed an unnamed intern for AI-fabricated citations was suspended from practice for two years, the panel finding that supervisory rules place responsibility on the signing lawyer regardless of who ran the prompt.⁷ In early 2026, a Nebraska lawyer whose divorce-appeal brief contained fifty-seven defective citations, including twenty wholly invented cases, was suspended indefinitely after initially denying, then admitting, that he had used AI; the state's highest court treated the denial itself, rather than the underlying error, as the more serious breach.⁸ Across these cases, a consistent pattern emerges: covering up the error draws a harsher penalty than the error itself.

India has produced the most consequential judicial response yet. In *Pooja Ramesh Singh v Jammu and Kashmir Bank Ltd*, the Supreme Court set aside orders of the National Company

³ *Mata v Avianca Inc* 2023 WL 4114965 (SDNY 22 June 2023).

⁴ 'AI Hallucination Cases: The 1,598-Case Sanctions Tracker' (HAQQ, 2026); 'AI Hallucination Sanctions Tracker: US Cases Where Courts Penalized AI Use' (Vaquill AI, 2026), discussing *Couvrette v Wisnovsky*, No 1:21-cv-00157-CL (D Or).

⁵ 'AI Hallucination Sanctions Tracker: US Cases Where Courts Penalized AI Use' (Vaquill AI, 2026), discussing *Whiting v City of Athens*.

⁶ 'AI Hallucination Cases: The 1,598-Case Sanctions Tracker' (HAQQ, 2026).

⁷ 'AI Hallucination Sanctions Tracker: US Cases Where Courts Penalized AI Use' (Vaquill AI, 2026), discussing the disciplinary proceedings against Zachariah Crabill.

⁸ 'AI Hallucination Legal Cases: A Sanctions Tracker' (GC AI, 2026), discussing the disciplinary proceedings against Greg Lake.

Law Tribunal and its appellate tribunal after finding that both had relied on precedents that were either wholly invented or real citations padded with fabricated paragraphs; the Court declared a 'zero-tolerance' policy toward unverified AI-generated precedent, held that citing such material is professional misconduct, and directed the Bar Council of India to frame disciplinary norms for the profession.⁹ The ruling followed a string of High Court decisions grappling with the same problem: the Bombay High Court imposed costs of ₹50,000 on a litigant whose submissions cited an untraceable judgment bearing the tell-tale hallmarks of AI output in *Deepak v Heart and Soul Entertainment Ltd*,¹⁰ while the Andhra Pradesh High Court took the more lenient view in *Gummadi Usha Rani v Sure Mallikarjuna Rao* that fabricated citations in an underlying order do not automatically void it a divergence the Supreme Court itself took up before delivering its own ruling five months later.¹¹ Earlier decisions, including the Punjab and Haryana High Court's warning in *Jaswinder Singh v State of Punjab* and the Manipur High Court's in *Md Zakir Hussain v State of Manipur*, show that Indian courts were flagging the problem well before it reached the Supreme Court.¹² Taken together, the trend mirrors what is visible in the United States and Singapore: courts moving toward institutional responses cost orders, bar referrals, and now a disciplinary framework while leaving the underlying duty exactly where it always was, on the lawyer who signs the filing.

BIAS BAKED INTO THE TRAINING DATA

Large language models learn patterns from the text on which they are trained, and that text carries the biases of the society that produced it. An AI tool asked to summarise 'typical' sentencing outcomes, assess the credibility of a witness statement, or predict case value can silently reproduce discriminatory patterns rather than correct for them. The problem is not new to AI generally, and the litigation history of algorithmic risk-assessment tools in criminal sentencing offers a preview of what awaits generative tools used for similar purposes.

In *State v. Loomis*, the Wisconsin Supreme Court upheld a trial court's reliance on a proprietary recidivism-risk score against a due-process challenge, while cautioning that such scores must

⁹ Pooja Ramesh Singh v Jammu and Kashmir Bank Ltd 2026 INSC 668 (2 July 2026); 'Landmark SC Ruling: AI Fake Precedents Void Judgments' (PrimeLegal, 3 July 2026).

¹⁰ *Deepak v Heart and Soul Entertainment Ltd* (Bombay HC, January 2026); '10 Cases That Show Indian Courts Have an AI Hallucination Problem' (MediaNama, 3 July 2026).

¹¹ *Gummadi Usha Rani v Sure Mallikarjuna Rao* (AP HC, January 2026); MediaNama (n 10).

¹² *Jaswinder Singh v State of Punjab*, CRM-M-22496-2022 (P&H HC, 27 March 2023); *Md Zakir Hussain v State of Manipur* (Manipur HC, 23 May 2024).

be used carefully and never treated as determinative of the sentence imposed.¹³ Subsequent empirical research into the same class of tool has found that individuals from marginalised groups are more likely to be misclassified as high risk despite comparable or less serious records, a pattern researchers attribute to the historical policing and sentencing data on which such tools are trained rather than to any intentional design choice.¹⁴ Because the underlying reasoning of a large language model is opaque in a similar way, a lawyer relying on its output may not recognise that bias has crept into an analysis until it surfaces in argument, or worse, in a judgment that adopts it uncritically.

Existing equal-protection doctrine, which generally requires proof of discriminatory intent rather than mere disparate impact, offers limited recourse once such a pattern has already influenced an outcome.¹⁵ That places the practical burden of catching the bias on the lawyer before the tool's output is relied upon at all. The European Union has taken a more preventive approach, classifying AI systems that assist judicial authorities in researching or applying the law as 'high-risk' under its AI Act, which triggers risk-management and human-oversight duties well beyond anything currently required of a lawyer's own use of comparable tools.¹⁶

CONFIDENTIALITY AND PRIVILEGE LEAKAGE

Pasting client facts, draft pleadings, or settlement strategy into a public AI chat interface risks two distinct harms: the information may be used to train future models, and the act of disclosure to a third-party system can itself be argued to waive privilege. The American Bar Association's Formal Opinion 512 addressed this directly, confirming that the existing duty of confidentiality under Model Rule 1.6, which protects all information relating to a client's representation regardless of its source, applies with undiminished force to generative AI tools, and that in some circumstances a client's informed consent is required before confidential information is entered into such a tool at all.¹⁷ The opinion also ties competence, under Model Rule 1.1, to an understanding of a tool's data-handling practices, since a consumer chatbot with no retention

¹³ *State v Loomis* 2016 WI 68, 881 NW 2d 749 (Wis 2016).

¹⁴ 'Beyond the Algorithm: Professional Experiences and Perceptions of AI Bias' (arXiv, 2026).

¹⁵ *Washington v Davis* 426 US 229 (1976).

¹⁶ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L 2024/1689, annex III.

¹⁷ American Bar Association Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512, 'Generative Artificial Intelligence Tools' (29 July 2024).

limits carries a materially different risk from an enterprise product with such protections built in.

Enterprise and so-called 'walled garden' AI products marketed to law firms address the retention concern to varying degrees, typically through contractual assurances that inputs will not be retained or used for training. But the underlying obligation does not change merely because the tool is convenient, and supervisory duties under Model Rule 5.3 extend the same responsibility to nonlawyer staff, meaning a firm cannot simply delegate the confidentiality risk downward.

OVERRELIANCE AND THE EROSION OF INDEPENDENT JUDGMENT

A subtler risk than outright fabrication is quiet overreliance: a lawyer who no longer independently reasons through a problem because the AI's answer feels sufficient. Formal Opinion 512 frames the required degree of independent verification as fact-specific, depending on the tool and task, but is clear that generative AI cannot substitute for a lawyer's competent legal work in any circumstance.¹⁸ Competence rules require lawyers to understand the tools they use well enough to recognise their limitations, not merely to operate them.

Treating AI output as a finished product rather than a first draft inverts that relationship, and the danger is greatest precisely where the output reads most fluently, since polished prose is easily mistaken for sound reasoning. Firms that introduce AI drafting tools without corresponding training on their limitations report that overreliance tends to increase, not decrease, over time as users grow more comfortable with the tool, unless the risk is actively managed through firm policy and periodic audits of AI-assisted work.

DUTY TO DISCLOSE AI USE TO COURTS AND OPPOSING COUNSEL

The disclosure landscape remains fragmented. There is no single nationwide rule in the United States requiring lawyers to disclose AI use in court filings. New York's statewide rule, 22 NYCRR Part 161, effective from 1 June 2026, permits AI use in preparing court papers and does not impose a general disclosure requirement, instead placing the emphasis on accountability for the resulting filing's accuracy regardless of how it was produced.¹⁹ The

¹⁸ Ibid.

¹⁹ New York Unified Court System, 22 NYCRR Part 161, 'Use of Artificial Intelligence Technology' (effective 1 June 2026); 'Disclosure Is Not Accountability: New York's AI Court Rule and the Future of Legal Responsibility' (AI Policy & Governance Institute, 2026).

Illinois Supreme Court's AI policy takes a similar accountability-over-disclosure posture.²⁰ Beneath these system-wide policies, however, hundreds of individual federal and state judges maintain standing orders that go considerably further: some require an affirmative certification with every brief or motion stating whether AI was used and confirming that a human independently verified every citation and factual assertion, a model associated with certain federal district judges in Texas and adopted in some form across dozens of federal districts.²¹

The result is that a lawyer's disclosure obligation can turn on which judge is assigned to the case rather than on any uniform national standard, and the only safe practice is to check the specific court's and judge's requirements before every filing. Whatever the formal disclosure rule in a given court, several disciplinary decisions have treated a lawyer's evasiveness about AI use, rather than the underlying citation error itself, as the aggravating factor in sanctions, as in the Nebraska case discussed above, where an initial denial of AI use converted a correctable drafting failure into grounds for indefinite suspension. Candour about tool use, even when it reveals a mistake, has consistently fared better before courts and disciplinary panels than concealment.

CONCLUSION

A practical checklist:

- Verify every citation, quotation, and case summary against the primary source before filing; treat AI output as an unverified first draft, never a final product.
- Confirm your specific court's and judge's AI disclosure requirements before every filing, since the obligation varies by jurisdiction and even by individual judge.
- Never enter client-identifying or privileged facts into a public, non-enterprise AI tool, and understand the data-retention practices of any enterprise tool you do use.
- Keep a record of what was AI-generated and what was independently checked, in case of a later challenge.
- Disclose AI use candidly if a court or rule requires it, and even where it does not, if asked directly; concealment has consistently drawn harsher sanctions than the underlying error.

²⁰ 'Court AI Disclosure Orders: Who Actually Requires What (2026)' (Legal AI Compliance Tracker, 2026).

²¹ 'AI Disclosure Requirements for Lawyers: What Courts Require in 2026' (Claude for Lawyers, 2026); 'When Must Lawyers Disclose AI Use? Court Requirements for Legal Work' (Spellbook, 2026).

- Treat supervisory sign-off as a genuine second check, not a formality: supervising lawyers have been sanctioned, and in some cases suspended, for filings they did not personally verify.

None of these counsels against using AI in legal practice; the efficiency gains are real, and the technology is not going away. What the case law makes clear, from New York to Wisconsin to New Delhi, is that the professional duties of competence, candour, and confidentiality apply with undiminished force regardless of which tool produced the first draft. As sanctions climb toward six-figure penalties, suspensions, and now a Supreme Court-mandated disciplinary framework, the practical margin for error keeps shrinking, and the lawyers faring best are consistently the ones who treat every AI output as a claim still to be checked, rather than a fact already to be trusted.