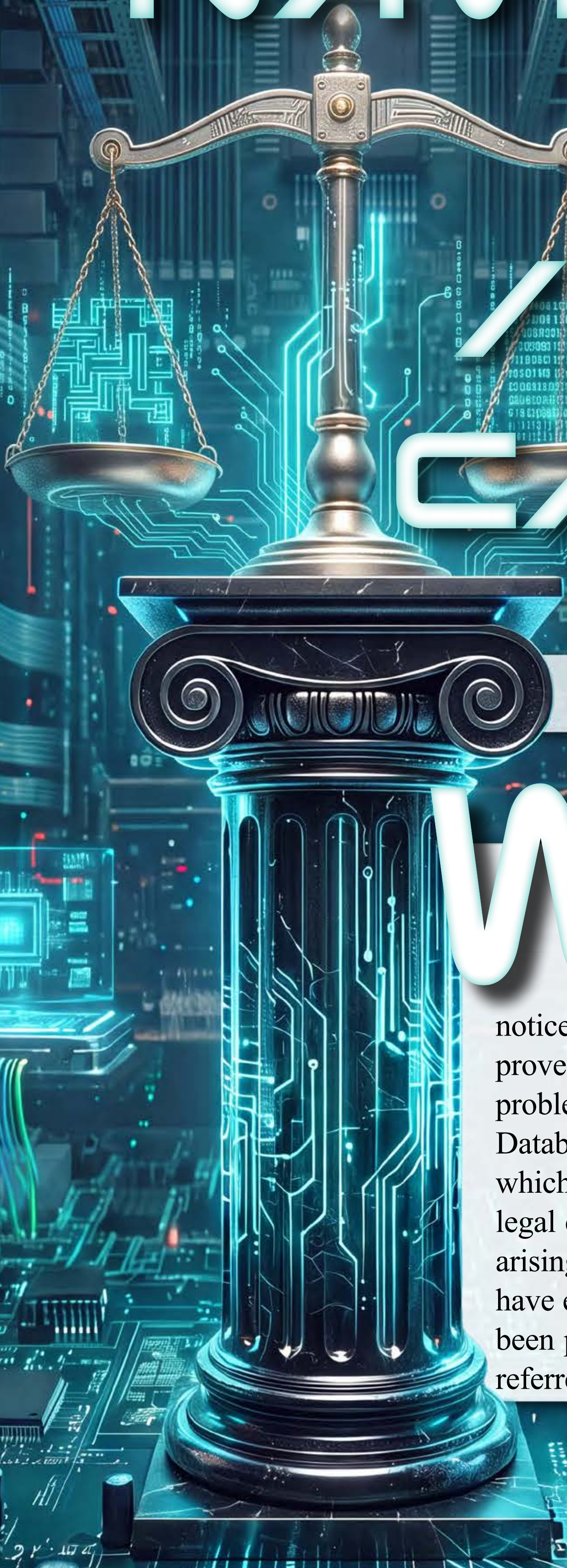
NAVIGATING THE AI RISK/BENEFIT CALCULUS



by Antonio (T.J.) Clayton II

When a New York federal court sanctioned two attorneys in *Mata v. Avianca, Inc.* for submitting a brief containing six fabricated case citations generated by ChatGPT,[1] the legal profession took notice. What many initially dismissed as an anomaly has since proved to be the leading edge of a significant and growing problem. Researcher Damien Charlotin's AI Hallucination Database now catalogs more than 1,200 cases worldwide in which generative artificial intelligence[2] produced fabricated legal content submitted to courts, with more than 700 of those arising in U.S. courts since January 2025 alone.[3] Sanctions have escalated well beyond monetary penalties; attorneys have been publicly reprimanded, disqualified from proceedings and referred to state licensing authorities.

The danger extends beyond fabricated citations. In April 2026, Anthropic announced that its Claude Mythos Preview model had demonstrated the ability to autonomously discover and exploit zero-day vulnerabilities in every major operating system and web browser.[4] The model identified thousands of previously unknown security flaws, prompting Anthropic to restrict its release to a small group of critical infrastructure partners under a coordinated initiative called Project Glasswing. The United Kingdom's AI Security Institute confirmed that Mythos was the first AI model to complete a simulated full-network takeover in its testing environment. For practitioners, the implications are significant: AI tools are no longer merely drafting assistants that might hallucinate a citation. They are rapidly becoming capable of penetrating the very systems on which law firms store confidential client data. The competence and confidentiality obligations that have always governed the profession now extend to understanding not just what an AI tool produces, but what AI systems are capable of doing to the digital infrastructure on which modern practice depends.

In a January 2024 letter to the Louisiana State Bar Association, the Court's General Counsel, writing on behalf of the Justices, observed that while AI applications in the legal profession are new, the ethical and professional rules "are not new and have been in place for decades" and remain broad enough to address the issues presented by AI.[5] The old rules, in short, still apply. This article surveys the national sanctions landscape, maps Louisiana's existing ethical framework to the risks posed by generative AI and offers practical guidance for practitioners navigating this rapidly evolving terrain.

The National Sanctions Landscape

The sanctions emerging from attorneys' misuse of generative AI have followed a clear trajectory.

Mata v. Avianca, Inc. remains the profession's most widely cited cautionary example. In that case, an attorney used ChatGPT to research a personal injury matter and submitted a brief containing six case citations that did not exist. When opposing counsel challenged the citations, the attorney returned to ChatGPT and asked the tool to confirm their authenticity — which it did, falsely. The court imposed a $5,000 sanction under Rule 11 and ordered the attorneys to notify every judge whose name had been falsely attributed to the fabricated opinions. The case illustrated what has since become the most common pattern in AI sanctions: an attorney relied on AI-generated output without any independent verification of its accuracy.

Two years later, *Johnson v. Dunn*[6] demonstrated how severely courts will respond when the same mistakes occur within a sophisticated practice environment. There, a partner and assistant practice group leader at Butler Snow LLP used ChatGPT to locate supporting caselaw and inserted fabricated citations into two discovery motions. An of-counsel attorney drafted, signed and filed both motions without independently reviewing the citations the partner had added. The supervising partner and practice group leader acknowledged that his ordinary practice was simply to assume others on the team had verified authorities. The court found that each attorney's conduct constituted recklessness tantamount to bad faith. The sanctions went well beyond a monetary penalty: the attorneys were publicly reprimanded, disqualified from further participation in the case and referred to the Alabama State Bar. Notably, Butler Snow had an internal policy requiring written approval before using ChatGPT for legal research, a policy the partner disregarded. The court characterized AI misuse as "a serious and time-sensitive problem" that, if left unchecked, would impose escalating costs on litigants and damage public confidence in the justice system.

The California Court of Appeal's decision in *Noland v. Land of the Free, L.P.*[7] introduced yet another dimension. There, plaintiff's appellate counsel submitted briefs in which nearly all legal quotations were fabricated: AI-generated text falsely attributed to real cases. The court sanctioned counsel $10,000 and ordered the opinion forwarded to the State Bar. But the decision may be most significant for the questions it raised about opposing counsel's obligations. Although the court declined to sanction respondents' attorneys, it noted that they had not alerted the court to the fabricated citations and appeared to have discovered the issue only when the court itself raised it. Whether this observation matures into an affirmative duty remains to be seen, but practitioners would be wise to take notice.

These decisions are not outliers. Courts have sanctioned attorneys for AI-generated fabrications in *Park v. Kim*,[8] where the 2nd Circuit referred counsel for discipline; in *Benjamin v. Costco Wholesale Corp.*,[9] resulting in a $1,000 penalty; in *Gauthier v. Goodyear Tire & Rubber Co.*,[10] imposing $2,000 in sanctions; and in *Coomer v. MyPillow, Inc.*,[11] where the court sanctioned each defense attorney $3,000. In *Wadsworth v. Walmart Inc.*,[12] a Wyoming federal court sanctioned two Morgan & Morgan attorneys and local counsel from Goody Law Group a total of $5,000 after the firm's internal AI platform generated eight nonexistent case citations in *motions in limine*; it also revoked the drafting attorney's *pro hac vice* admission. The pattern is unmistakable.

These risks are no longer hypothetical for Louisiana practitioners. In *Woodward Harbor L.L.C. v. City of Mandeville*, the Eastern District of Louisiana confronted an opposition brief that plaintiffs' counsel had prepared using ChatGPT and a commercial AI research tool, and that was, in the court's words, "replete with erroneous and hallucinated citations to non-existent caselaw." In granting the City's motion to dismiss, the court found that the brief repeatedly relied on cases that do not exist, invented quotations and mischaracterized holdings, attributing the errors to "an astonishingly careless use of generative AI." The court stressed that the brief had been signed by attorneys from two Louisiana firms and called their failure to perform even a cursory check of their authorities "wholly unacceptable," echoing the supervisory breakdown that drove the sanctions in *Johnson*. Pointedly, opposing counsel had already flagged the nonexistent cases in its reply brief, yet plaintiffs' counsel never moved to correct them. Following a separate show-cause proceeding, the court imposed a $1,000 sanction on the attorney responsible and ordered him to complete continuing education on generative AI, while declining to sanction co-counsel who neither used nor knew of the tool. For Louisiana lawyers, *Woodward Harbor* confirms that the national trend has arrived in-state, and that the State's existing rules, from Rule 1.1's competence requirement to the certification compelled by Code of Civil Procedure article 863, will be enforced here as rigorously as elsewhere.

Louisiana's Ethical Framework

In January 2024, the Louisiana Supreme Court addressed the profession's obligations in a letter from General Counsel David Becker to then-LSBA President Shayna Sonnier.[13] Writing on behalf of the Justices, Becker acknowledged that AI presents both opportunities and concerns from an ethics and professionalism standpoint. But the letter's central message was unambiguous: while AI applications in the legal profession may be new, the ethical rules governing the bench and the Bar "are not new and have been in place for decades" and remain "robust and broad enough to cover the landscape of issues presented by AI in its current forms." The Court specifically identified Rules of Professional Conduct 1.1, 1.3, 1.4, 1.6, 1.15, 3.1, 3.3 and 5.3, along with Code of Civil Procedure articles 371 and 863, as the governing framework. The Court's Technology Commission and Rules Committee continue to monitor developments, but the message to practitioners is clear: the existing rules apply now, in full.

Those rules apply directly to the risks illustrated by the national sanctions cases. Rule 1.1 requires competent representation, which the LSBA has already interpreted to encompass technological competence. In Public Opinion 19-RPCC-021,[14] issued in 2019, well before generative AI entered widespread legal use, the LSBA concluded that lawyers using technology must do so competently and diligently, and that when relying on nonlawyer technology providers, Rule 5.3 requires reasonable efforts to ensure the provider's conduct is compatible with the lawyer's professional obligations. That framework extends naturally to AI: a lawyer who uses a generative AI tool without understanding its propensity to fabricate legal authority has failed the competence standard, just as the attorneys in *Mata* and *Johnson* did.

Rule 1.6, as amended in Louisiana in 2015 to add subsection (c), requires lawyers to make reasonable efforts to prevent the inadvertent or unauthorized disclosure of client information. Inputting confidential client data into a consumer-grade AI tool, where the provider may retain, train on or otherwise process that data, raises serious questions under this rule.[15] Rule 3.3's duty of candor toward the tribunal is implicated every time an attorney presents AI-generated citations without verification, as the courts in *Mata* and *Noland* recognized. In *Johnson*, the court stopped short of applying Rule 3.3, noting that the rule by its terms forbids only knowing misstatements and that the attorneys' failures were reckless rather than intentional; but the court nonetheless observed that submitting unverified AI-generated citations is "wholly inconsistent with the duty of candor." Whether or not Rule 3.3 reaches reckless conduct in every jurisdiction, the practical lesson is the same: verification is the practitioner's obligation, not the tool's. And Rules 5.1 and 5.3, governing supervisory responsibilities, take on particular significance when AI is understood not as a trusted colleague but as a tool requiring the same oversight one would apply to a junior associate or paralegal, the very framework the ABA adopted in Formal Opinion 512.[16]

Louisiana Code of Civil Procedure article 863 reinforces these obligations from a procedural standpoint. An attorney's signature on a pleading certifies that the document is well-grounded in fact and warranted by existing law. That certification is not diminished because the document was drafted with AI assistance. As the 5th Circuit recognized when it declined to adopt a standalone AI disclosure rule in June 2024,[17] the existing obligations already require that filings "be carefully checked for truthfulness and accuracy." The court added a pointed warning that has particular resonance for practitioners in this circuit: "'I used AI' will not be an excuse for an otherwise sanctionable offense."

Practical Guidance

The ethical framework is clear. The practical question is how Louisiana practitioners should operationalize it. Several principles, drawn from national guidance and reinforced by the cases discussed above, provide a workable starting point.

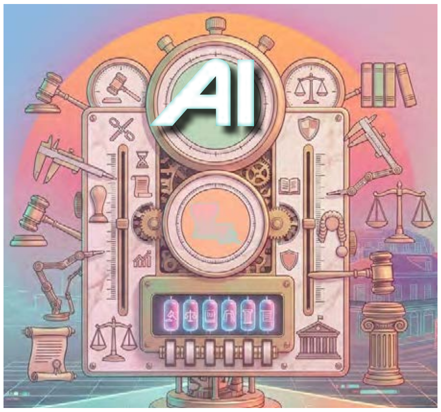
Generative AI should be understood through the lens of existing supervisory obligations. ABA Formal Opinion 512, issued in July 2024, provides the most comprehensive national guidance to date on lawyers' use of generative AI tools. The opinion addresses competence, confidentiality, communication, candor, supervisory responsibilities and fees, organizing each under the applicable Model Rules. Under Model Rules 5.1 and 5.3, managerial and supervisory lawyers must establish clear policies regarding AI use, provide training on associated risks and ensure that all AI-assisted work product is reviewed before submission. The functional result is that AI-generated work product warrants the same level of scrutiny a lawyer would apply to a memorandum prepared by a paralegal or junior associate. The tool may draft; the lawyer must verify.

Firm AI policies must be implemented and enforced — not merely adopted.[18] *Johnson v. Dunn* is the starkest illustration of this point. Butler Snow had a policy requiring written approval before using ChatGPT for legal research. The partner who generated the fabricated citations disregarded it. Although the court released the firm without sanction after finding it had acted reasonably in its preventive efforts, the individual attorneys were publicly reprimanded, disqualified and referred to the state bar. A policy that exists on paper but is not reinforced through training, supervision and accountability offers little protection when sanctions are at issue. Effective policies should identify approved AI tools, require independent verification of all AI-generated content, prohibit the input of confidential client information into consumer-grade platforms and mandate periodic training for all attorneys and staff. In light of the rapidly evolving cybersecurity landscape, those policies should also address the security architecture of approved AI platforms, establish protocols for evaluating new tools before deployment and include incident response procedures in the event of a data breach involving AI-processed client information.

Confidentiality demands particular attention, and the risks extend beyond ethical obligations to discoverability and security. Louisiana Rule 1.6(c) requires reasonable efforts to prevent the inadvertent or unauthorized disclosure of client information. When a lawyer inputs client-specific data into a third-party AI tool, that information may be retained, used for model training or otherwise processed in ways the lawyer cannot control. Florida Bar Opinion 24-1,[19] issued in January 2024, recommended that lawyers obtain the affected client's informed consent before using a third-party generative AI program if doing so would involve the disclosure of confidential information. Although that opinion is specific to Florida, its reasoning applies with equal force under Louisiana's parallel rule. The distinction between enterprise AI platforms, which typically offer contractual data protections, and consumer-grade tools is a material one that practitioners should evaluate before incorporating any AI tool into client work.

The AI platform distinction also carries consequences beyond ethics compliance. Recent federal decisions have addressed whether a party's communications with AI tools are discoverable, and the choice of platform and its terms of service figure prominently in the analysis. In *Warner v. Gilbarco Inc.*,[20] the Eastern District of Michigan held that a *pro se* litigant's interactions with ChatGPT for purposes of formulating legal strategy were entitled to work-product protection under Federal Rule of Civil Procedure 26(b)(3), which extends protection to materials prepared by parties, not merely attorneys. The court rejected the argument that disclosing information to an AI tool waives that protection, reasoning that an AI tool is not an adversary and that ordering disclosure would nullify work-product protection wherever software is used as a drafting aid. The District of Colorado reached the same conclusion in *Morgan v. V2X Inc.*,[21] following *Warner*'s reasoning while holding that the identity of the specific AI tool used was not itself protected. Although *United States v. Heppner*[22] — a criminal case in which the Southern District of New York found that a defendant's communications with an AI tool were neither privileged nor work-product protected — may appear contrary, that court explicitly noted that work-product protection under the federal discovery rules was not at issue because the government obtained the materials through a search warrant, not through pretrial discovery. Critically, the *Heppner* court also found that the defendant lacked a reasonable expectation of confidentiality because the AI tool's terms of service permitted the provider to collect and disclose user data. The emerging framework in the civil discovery context confirms what the ethical rules already require: platform choice matters, terms of service matter and the existing rules already provide the governing structure.

The cybersecurity dimension of con-



fidentiality has likewise taken on new urgency. The emergence of AI models with autonomous vulnerability-discovery capabilities means that the systems on which law firms rely, including cloud-based AI platforms, may themselves become targets of increasingly sophisticated cyberattacks. Rule 1.6(c)'s requirement of "reasonable efforts" to prevent unauthorized disclosure now encompasses not only understanding how an AI provider handles client data but also evaluating the security architecture of the platforms through which that data is transmitted and stored. Firm AI policies should address cybersecurity protocols alongside confidentiality and verification requirements, including encryption standards, vendor security assessments and periodic reviews of approved AI tools.

Independent verification of AI output is nonnegotiable. Every citation, every factual assertion and every legal conclusion generated by AI must be confirmed before it is presented to a court or relied upon in advising a client.[23] The lesson of *Mata*, where an attorney asked ChatGPT to confirm the validity of its own fabricated citations, bears repeating: the tool that generated the output cannot be the tool that verifies it. This principle is not novel; it is the same duty of care that has always required lawyers to confirm the accuracy of their work before submitting it, regardless of how that work was produced.

Billing practices must also reflect the reality of AI-assisted work. Louisiana Rule 1.5 requires that fees be reasonable. Florida Opinion 24-1 addressed this issue directly, concluding that while a lawyer may charge for reasonable time spent on case-specific research and drafting when using generative AI, the increased efficiency "must not result in falsely inflated claims of time." ABA Formal Opinion 512 similarly noted that a fee charged for work in which little or no effort was performed is unreasonable regardless of the billing structure. Practitioners should communicate transparently with clients about AI use and consider whether alternative fee arrangements better align the benefits of AI efficiency with the client's interests.[24]

Conclusion

The national sanctions landscape leaves little room for doubt: courts will hold attorneys accountable for AI-generated errors with the same rigor they apply to any other failure of professional responsibility and, increasingly, with greater severity. From the $5,000 fine in *Mata v. Avianca* to the disqualification and bar referrals in *Johnson v. Dunn* to the modest $1,000 fine but burdensome CLE imposed in *Woodward Harbor v. City of Mandeville*, the trajectory is unmistakable.

What remains is execution. The rules that Louisiana practitioners have followed throughout their careers are more than sufficient to navigate this new terrain, provided those practitioners apply the same skepticism and diligence to AI-generated output that they would to the work of any unsupervised assistant. That means verifying every citation, safeguarding every confidence and billing with transparency. The technology is new. The professional obligations are not. And the imperative extends beyond avoiding sanctions. Generative AI represents one of the most significant shifts in the practice of law in a generation, and practitioners who fail to engage with it responsibly will find themselves at a growing disadvantage to those who do. The old rules still apply, and for Louisiana lawyers prepared to follow them, this moment is less of a threat than an opportunity.

FOOTNOTES

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

2. Generative artificial intelligence refers to AI systems that produce new text, images, or other content in response to user prompts. These systems—including large language models such as OpenAI's ChatGPT, Google's Gemini, Anthropic's Claude, and Microsoft's Copilot—are trained on large datasets and generate outputs by predicting likely sequences of words. They are distinct from earlier forms of AI used in legal practice, such as predictive coding in e-discovery or contract analysis tools, which classify or extract information from existing documents rather than generating new content.

3. Damien Charlotin, AI Hallucination Database, https://www.damiencharlotin.com/hallucinations/ (tracking documented instances of AI-generated fabrications in court filings worldwide).

4. Nicholas Carlini et al., *Assessing Claude Mythos Preview's Cybersecurity Capabilities*, Anthropic (Apr. 7, 2026), https://red.anthropic.com/2026/mythos-preview/. Anthropic restricted the model's release to a limited group of critical infrastructure partners, including Amazon Web Services, Apple, Google, JPMorganChase, Microsoft, and Nvidia, under an initiative designated Project Glasswing.

5. Letter from David Becker, Gen. Counsel, La. Supreme Court, to Shayna Sonnier, President, La. State Bar Ass'n (Jan. 22, 2024). The letter and its implications are discussed in detail in Part II, infra.

6. *Johnson v. Dunn*, 792 F. Supp. 3d 1241 (N.D. Ala. 2025).

7. *Noland v. Land of the Free, L.P.*, 114 Cal. App. 5th 426, 336 Cal. Rptr. 3d 897 (2025).

8. *Park v. Kim*, 91 F.4th 610 (2d Cir. 2024).

9. *Benjamin v. Costco Wholesale Corp.*, 779 F. Supp. 3d 341 (E.D.N.Y. 2025).

10. *Gauthier v. Goodyear Tire & Rubber Co.*, No. 1:23-CV-281, 2024 WL 4882651 (E.D. Tex. Nov. 25, 2024).

11. *Coomer v. MyPillow, Inc.*, No. 1:22-cv-01129-NYW-SBP (D. Colo. July 7, 2025).

12. *Wadsworth v. Walmart Inc.*, 348 F.R.D. 489, No. 23-CV-118, 2025 WL 608073 (D. Wyo. Feb. 24, 2025).

13. Letter from David Becker, Gen. Counsel, La. Supreme Court, to Shayna Sonnier, President, La. State Bar Ass'n (Jan. 22, 2024).

14. La. State Bar Ass'n, Pub. Op. 19-RPCC-021 (Feb. 6, 2019).

15. The distinction between "open" and "closed" AI systems is material to the confidentiality analysis. Consumer-grade tools such as ChatGPT's free tier may retain user inputs for model training, meaning client data entered into the system could be incorporated into the model's training dataset and potentially surfaced in responses to other users. Enterprise platforms, by contrast, typically offer contractual commitments that user data will not be used for training and will be processed in accordance with specified data-handling protocols. Practitioners should review the terms of service and data processing agreements of any AI tool before inputting client information. For a recent commentary arguing that state bars should move beyond general reasonableness standards and evaluate specific AI tools by plan tier, see Nick Berk, *State Bars Need To Get Specific About AI Confidentiality*, Law360 (Apr. 13, 2026).

16. ABA Comm. on Ethics & Prof'l Responsibility, Formal Op. 512 (July 29, 2024).

17. U.S. Court of Appeals for the Fifth Circuit, Notice Regarding Use of Artificial Intelligence (June 11, 2024).

18. ABA Formal Opinion 512 concludes that supervisory obligations under Model Rules 5.1 and 5.3 require firms to establish policies governing the use of generative AI tools. Practitioners developing or updating firm AI policies may wish to consult ethics counsel to ensure those policies address the specific risks and obligations applicable to their practice.

19. Fla. State Bar Ass'n, Prof'l Ethics Comm., Op. 24-1 (Jan. 19, 2024).

20. *Warner v. Gilbarco Inc*., No. 2:24-cv-12333, 2026 WL 373043, at *4 (E.D. Mich. Feb. 10, 2026).

21. *Morgan v. V2X Inc*., No. 25-cv-01991-SKC-MDB, 2026 WL 864223, at *4–5 (D. Colo. Mar. 30, 2026).

22. *United States v. Heppner*, ___ F. Supp. 3d ___, No. 25 Cr. 503 (JSR), 2026 WL 436479 (S.D.N.Y. Feb. 17, 2026). For a comprehensive analysis reconciling these decisions, see Joshua Dunn, *2 Discovery Rulings Break with Heppner on AI Privilege Issue*, Law360 (Apr. 17, 2026).

23. Verification requires confirming that each cited authority exists, that the reporter citation is accurate, and that the case stands for the proposition cited. This should be done through an authenticated legal research platform such as Westlaw or Lexis, not through the AI tool that generated the citation.

24. Whether an attorney must disclose AI use to a client depends on the circumstances. ABA Formal Opinion 512 addresses this under Model Rule 1.4 (Communication), noting that lawyers should keep clients reasonably informed about the means by which their matter is being handled. Where AI use materially affects the scope, cost, or nature of the representation, disclosure and discussion with the client are warranted. Florida Opinion 24-1 goes further, recommending informed consent before using a third-party generative AI tool if doing so involves the disclosure of confidential information.

Antonio (T.J.) Clayton II is a rising second-year student at the University of Texas School of Law. Before law school, he earned a Master of Management Studies from Duke University's Fuqua School of Business and worked in equity research and venture capital. He is currently a summer associate at Arnold & Itkin in Houston. He can be reached at tjclayton@my.utexas.edu. Austin, Texas.