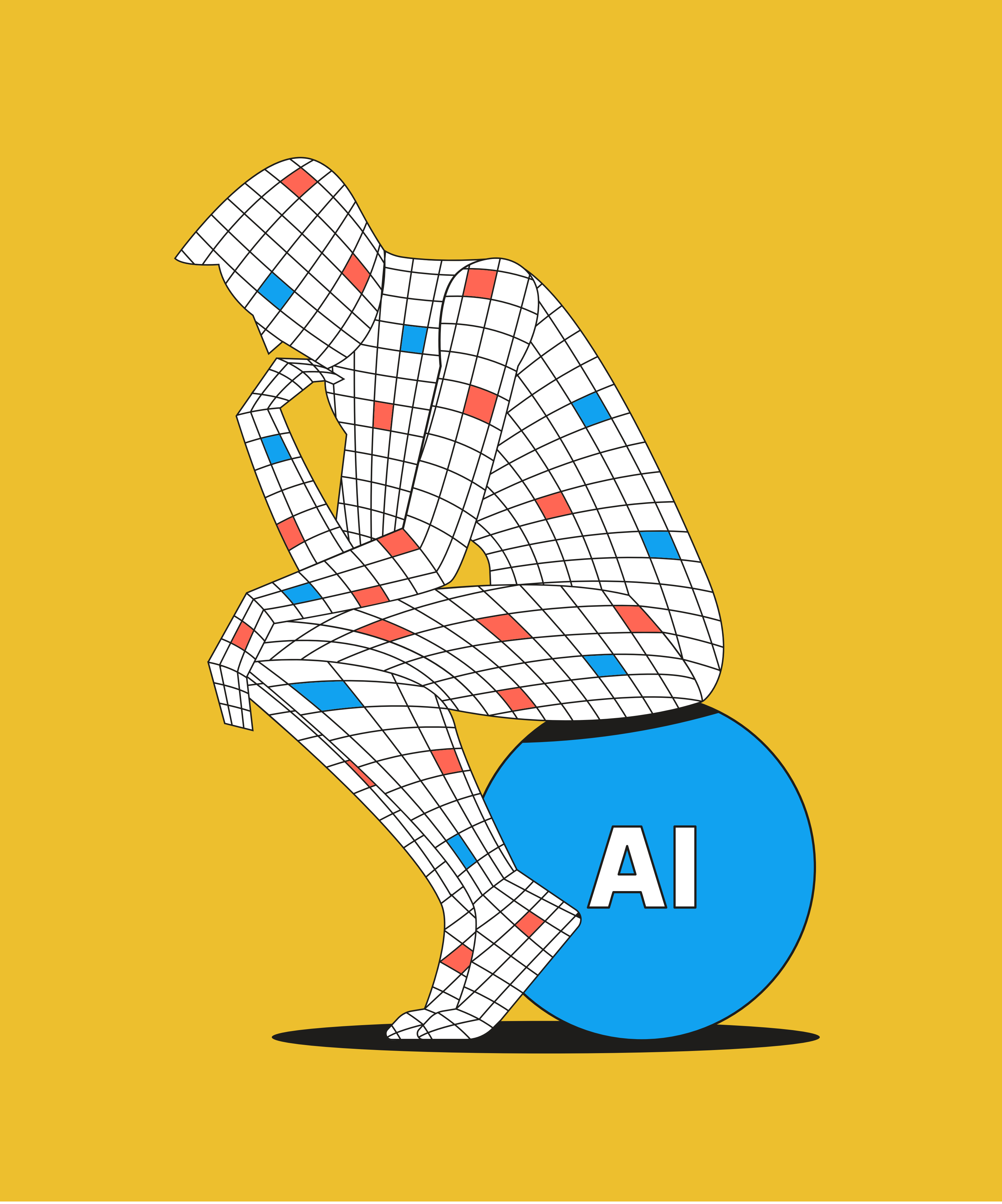

AI

ARTIFICIAL ETHICS?

ABA Formal Opinion 512 and the Role of Ethics in Artificial Intelligence

By J. Cliff McKinney II

For decades, books, movies, and television have been obsessed with artificial intelligence turning on humanity and destroying us. From the Thinking Machines in *Dune* to Skynet in *The Terminator* to the Cylons in *Battlestar Galactica*, humanity rarely survives once machines become intelligent. The threat is usually imagined as physical, taking the form of killer robots or systems that seize control. But for lawyers, at least for now, the more immediate risk is not physical; it is ethical.

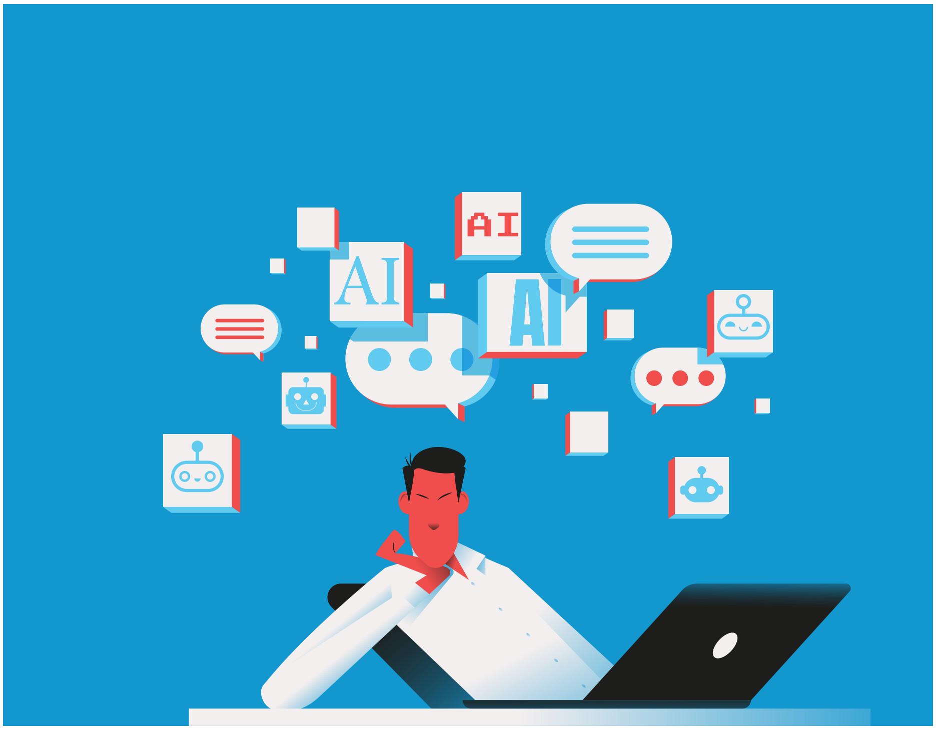
AI is rapidly changing the practice of law. It is embedding itself into the systems lawyers already use every day, including Microsoft Office, Westlaw, Lexis, and others. Lawyers are already using AI tools to draft documents, analyze issues, brainstorm strategies, and test arguments. Lawyers must learn important technical skills to use AI effectively, such as prompt engineering. But lawyers also must learn to use AI ethically and in compliance with the Rules of Professional Responsibility.

Courts are already confronting a growing number of cases involving lawyers who misuse AI. The most common misuse involves hallucinated citations, which are cases or authorities that do not exist but are presented as law. Other issues are beginning to emerge as well, including failures in supervision, risks to confidentiality, client communication, and questions about billing practices.

In July 2024, the American Bar Association (ABA) issued Formal Opinion 512 to address these issues. The Opinion takes a measured approach, applying the existing Rules of Professional Responsibility to the use of AI instead of proposing new rules. Now that significant caselaw is developing, it is clear that Opinion 512 is correct in most respects. But the ABA issued the Opinion at an early stage in the development of AI, before courts had meaningfully addressed how these issues would unfold in practice. For a much more detailed analysis of Formal Opinion 512, the developing case law, and ethical challenges, see my series in the *Arkansas Law Review: Ethics of Artificial Intelligence for Lawyers* (six-part series), Arkansas Law Notes (2026). The Arkansas Law Review series also includes a tool kit for lawyers, including sample engagement letter language, a model law firm policy on the use of AI, and a model training plan for legal professionals.

Getty Images

J. Cliff McKinney II is a managing member at Quattlebaum, Grooms & Tull PLLC in Little Rock, Arkansas. He is the Technology Vice-Chair of the Section's Joint Technology and Law Practice Group.



This article uses Formal Opinion 512 as its framework but focuses on how ethical situations are actually emerging. It examines how courts are applying the Rules of Professional Conduct to AI, how the law is developing inconsistently, and what lawyers must do now to avoid ethical pitfalls.

Before proceeding further, it is important to define what is now commonly meant by "AI." Several types of artificial intelligence have existed in computer science for some time. But the legal field is being transformed by one relatively new type of AI, known as generative artificial intelligence. Generative AI enables computers to produce highly accurate, detailed responses to queries. The technology relies on pattern recognition and the ability to generate new content. Simply put, generative AI systems have learned from trillions of pieces of information and can use patterns derived from those data to "guess" how to respond. Generative AI systems usually guess correctly, but there is no way to know when they are wrong without verifying the output. This weakness, along with AI's innate desire to train on new data, leads to many of the ethical issues this article examines.

Formal Opinion 512: The Framework—and Its Limits

Formal Opinion 512 is the ABA's first comprehensive attempt to address the ethical use of AI by lawyers. The Opinion focuses on six key areas: competence (Rule 1.1), confidentiality (Rules 1.6, 1.9(c), and 1.18(b)), communication (Rule 1.4), candor to the tribunal (Rule 3.3, together with Rules 3.1 and 8.4(c)), supervision (Rules 5.1 and 5.3), and fees (Rule 1.5). The Opinion emphasizes that lawyers must understand the capabilities and limitations of AI to satisfy Rule 1.1, protect client information consistent with Rule 1.6, supervise its use under Rules 5.1 and 5.3, and verify the accuracy of any output before relying on it to avoid violating Rules 3.1 and 3.3. The Opinion's central premise is that AI is a tool, and lawyers remain responsible for the work produced with that tool. The duties that have always governed the practice of law continue to apply, and there is nothing so magical about AI that it changes the fundamental rules.

But the Opinion also reflects the moment in which it was written. Formal Opinion 512 was issued at a time when the risks of AI were still largely theoretical. The Opinion anticipated problems, but it did not yet have the benefit of a developed body of caselaw. Two gaps have already become apparent.

First, the Opinion understates the role that Rule 3.3, candor to the tribunal, is playing in the emerging caselaw. Although the Opinion groups Rule 3.3 together with Rules 3.1 and 8.4(c), courts are treating candor as a central issue. Judges are not applying the rules uniformly. Some courts treat the submission of hallucinated authority as a violation of Rule 3.3 regardless of intent. Others focus on whether the lawyer "knowingly" made a false statement. The result is a developing split in the caselaw applying Rule 3.3, a divide that the Opinion does not anticipate.

Second, the Opinion largely overlooks candor outside the courtroom. Rule 4.1 governs truthfulness in statements to third persons, yet the Opinion does not address this rule. That omission matters in practice. Lawyers use AI not only in court filings, but in demand letters, negotiations, and other communications where accuracy still matters. A hallucinated case used to support a settlement demand raises the same concerns under Rule 4.1 as a false citation in a brief does under Rule 3.3. In some respects, the risk may be more significant because it is less likely to be immediately challenged and may be more difficult to prove.

Competence (Rule 1.1): Understanding AI as a System, Not a Source

Appropriately, Formal Opinion 512 begins with competence. The duty of competence under Rule 1.1 has long required lawyers to maintain not only legal knowledge, but also an understanding of the tools used in practice. The Opinion speculates that attorneys may soon have an ethical duty to understand how generative AI systems operate in order to remain competent. Some states, like Florida, already have added to their Rule 1.1 commentary a requirement to understand AI.

To use AI competently, lawyers must understand that generative AI creates new content rather than merely retrieving verified, human-created information. It produces responses based on patterns and probabilities, which inherently leads to errors. Comprehending that distinction alone helps lawyers avoid many of the risks associated with AI, including hallucinations and misplaced confidence in well-written output.

But in practice, competence requires more than recognizing that AI can be wrong. It requires understanding *how* and *when* it is likely to be wrong, as well as how to use AI to mitigate these risks. For example, AI performs well when asked to organize information, generate initial drafts, create summaries, or identify general issues. It performs less reliably when asked to provide precise legal authority, apply jurisdiction-specific rules, make legal judgments, or evaluate fact-intensive scenarios.

This becomes particularly important in real estate and probate practice. A lawyer reviewing a lease with AI may receive a summary that omits a key co-tenancy provision or mischaracterizes an assignment restriction. A probate attorney may receive trust language that appears balanced but introduces ambiguity in distribution standards or creates an unintended tax consequence. In both cases, the output is usually not obviously wrong and may be almost entirely accurate—*almost*. Competence requires the lawyer to recognize and carefully verify AI outputs to correct these errors.

Confidentiality (Rules 1.6, 1.9(c), and 1.18(b)): The Silent Risk

Lawyers always must remember that AI craves new information and can absorb everything it encounters. This poses a substantial ethical risk whenever a lawyer uses AI, as the AI might learn confidential information that is then exposed to other users. There are, however, ways to ethically manage this risk.

The simplest solution is to use "deidentification" or "anonymization" techniques. This means purging confidential information before using AI tools. For example, when using AI to review a purchase agreement, the attorney might change the buyer's name

Many of the AI products marketed to lawyers are "wrappers," meaning they rely on third-party large language models such as ChatGPT, Gemini, or Claude.

to X, the seller's name to Y, the purchase price to $1, and the location to 123 Street, Anytown, Texas. As another example, a probate attorney evaluating a trust structure could describe relationships in general terms rather than naming specific individuals. This prevents the AI from absorbing any confidential information.

Another tactic is to use AI tools that offer confidentiality protection. Many paid AI systems allow users to limit or disable the use of their data for model training and provide controls over data retention, although the scope of those protections varies by provider and product. Some providers also maintain recognized privacy and information management standards, such as ISO/IEC 27701, or undergo independent control audits, such as SOC 2. These frameworks and reports do not guarantee confidentiality, but they provide evidence that the vendor has implemented and tested structured controls designed to protect data.

It is also important to remember that many of the AI products marketed to lawyers are "wrappers," meaning they rely on third-party large language models such as ChatGPT, Gemini, or Claude. As a result, using these tools may involve transmitting data not only to the vendor but also to one or more underlying model providers. The key is to ensure that any such data flows are governed by contractual protections, particularly regarding model training, data retention, and downstream use.

Some firms have attempted to minimize confidentiality risks by developing in-house, proprietary AI systems. While this approach may reduce reliance on external vendors, it introduces a different set of ethical and operational risks that must be carefully managed. Firms must ensure that these systems are designed with access controls that respect ethical walls and other client-specific confidentiality restrictions. A system that aggregates or indexes firmwide data without appropriate segmentation could inadvertently expose information to lawyers who are prohibited from accessing it. Contractual data retention and deletion obligations present an equally serious challenge. Firms routinely receive large volumes of due diligence materials subject to contractual obligations requiring deletion if a transaction does not close, yet once those materials have been ingested by an AI system, complete removal may be difficult, particularly because many systems are not designed to fully "unlearn" information that has already been incorporated into model outputs. Systems that store data in an external retrieval index rather than embedding them in model weights can make deletion more manageable but do not eliminate the obligation to confirm that all components of the data pipeline have been addressed. In addition, some proprietary systems continue to rely on third-party large language models for processing, which means sensitive data may still be transmitted outside the firm's internal environment.

Candor to the Tribunal (Rule 3.3, Together with Rules 3.1 and 8.4(c)): Where Formal Opinion 512 Understates the Risk

Formal Opinion 512 addresses the duty of candor to the tribunal under Model Rule 3.3, but this is the main area where there is a gap between the Opinion and how the issues are playing out in court. The Opinion states that lawyers must not make false statements of law or fact to a court and must verify the accuracy of any AI-generated content before relying on it. The duty of candor always has required lawyers to ensure that their filings are accurate and supported. But AI-hallucinated authorities, ranging from fictitious quotes to nonexistent cases, are creating new challenges for courts, and courts are split about how to apply Rule 3.3.

Formal Opinion 512 treats Rule 3.3 as part of a broader group of rules dealing with candor to the tribunal, discussing it alongside Rules 3.1 and 8.4(c) in a relatively brief section. In practice, however, Rule 3.3 has become the focal point of AI-related sanctions cases, and courts cite it in approximately 80% of AI-related sanctions cases. This rule is also where courts are beginning to diverge in significant ways.

Two distinct approaches are emerging. The first is a strict view under which the submission of hallucinated authority is per se a violation of Rule 3.3, regardless of the lawyer's intent. Courts applying this approach treat the act of presenting false authority as incompatible with the duty of candor, even if the attorney did not realize the citations were hallucinated. In *United States v. Hayes*, 763 F. Supp. 3d 1054 (E.D. Cal. 2025), for example, the court devoted substantial attention to Rule 3.3 and emphasized that lawyers have an affirmative obligation to verify the accuracy of what they submit. The failure to do so, whether due to reliance on AI or otherwise, was sufficient to justify sanctions. Similarly, in *ByoPlanet Int'l, LLC v. Johansson*, 2025 WL 2091025 (S.D. Fla. July 17, 2025), the court characterized reliance on hallucinated authority as "abusive" and "bad-faith conduct," imposing significant sanctions. These cases reflect a view that the duty of candor imposes an absolute obligation to ensure accuracy, and that failure to verify is itself a violation of Rule 3.3.

The second approach focuses on the text of Rule 3.3, leading to a *mens rea* requirement. Model Rule 3.3 prohibits "knowingly" false statements of law or fact. In *Garner v. Kadince, Inc.*, 571 P.3d 812 (Utah Ct. App. 2025), the court concluded that the attorney's conduct—submitting hallucinated cases prepared by a law clerk—did not rise to the level of a Rule 3.3 violation because the attorney did not knowingly make a false statement, even though sanctions were still imposed on other grounds. The same reasoning appears in *Johnson v. Dunn*, 2025 WL 2086116 (N.D. Ala. July 23, 2025), where multiple attorneys signed a pleading containing hallucinated citations. The court described the conduct as "wholly inconsistent with the duty of candor," but ultimately declined to find a Rule 3.3 violation because the false statements were not made knowingly, though the court still imposed significant Rule 11 sanctions.

Supervision (Rules 5.1 and 5.3): The Responsibility Does Not Move

Formal Opinion 512 discusses supervision under Model Rules 5.1 and 5.3, noting that its practical effects extend beyond what the Opinion originally outlines. Lawyers must oversee both subordinate lawyers and nonlawyer assistants to ensure everyone follows the Rules of Professional Conduct. When AI is misused, responsibility does not stay with the person who typed the prompt. In *Wadsworth v. Walmart Inc.*, 348 F.R.D. 489 (D. Wyo. 2025), motions containing nine citations—eight of which were hallucinated—were prepared by pro hac vice counsel using an AI tool. The court did not limit discipline to the drafting attorney. The court sanctioned the pro hac vice attorney and revoked his admission. But the court also fined the lead counsel and the local counsel who had sponsored the pro hac vice admission, even though neither of them prepared the filing.

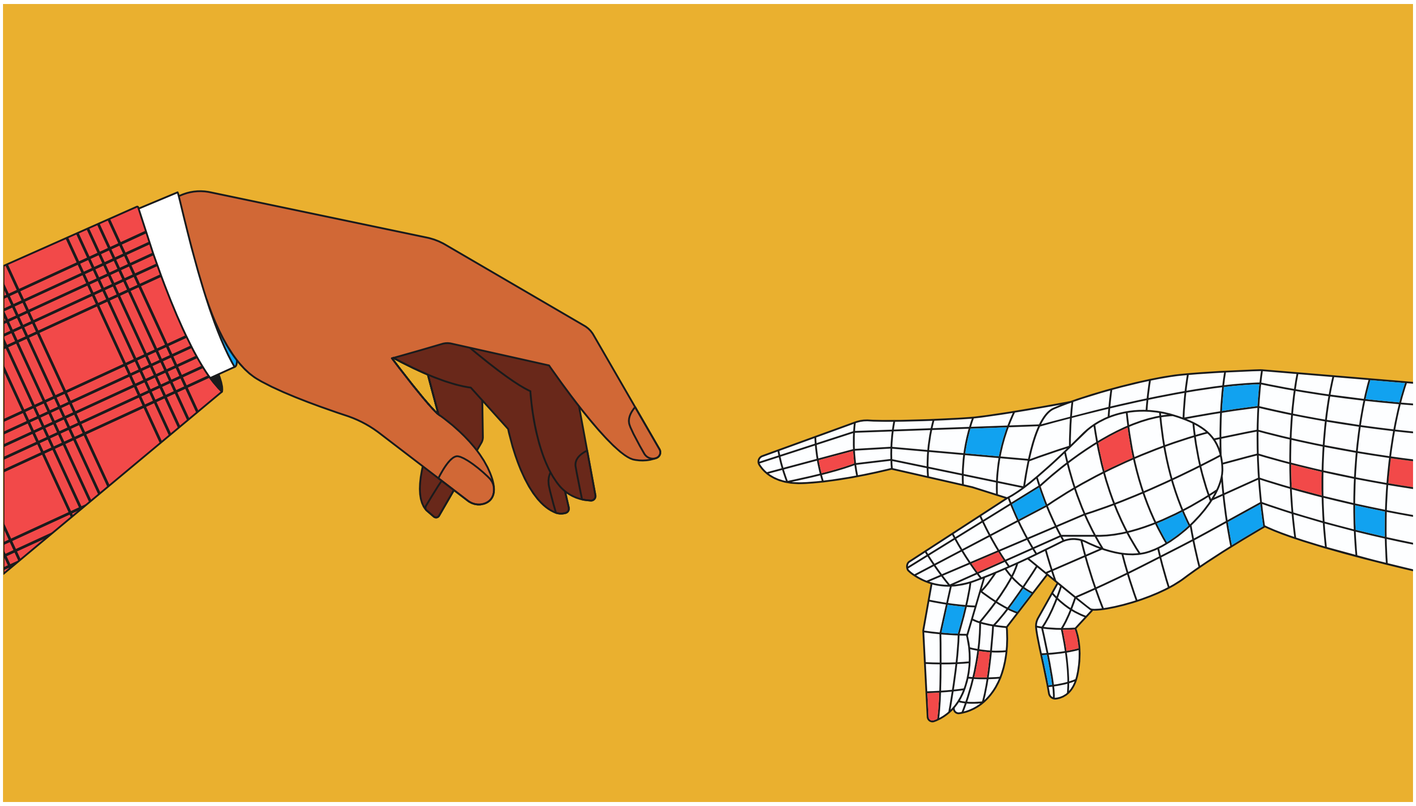
The same principle appears in *Versant Funding LLC v. Teras Breakbulk Ocean Navigation Enters., LLC*, 2025 WL 1440351 (S.D. Fla. May 20, 2025). There, a hallucinated citation remained in a filing even after opposing counsel flagged it. The court emphasized not only the underlying error, but the delay in correcting it, characterizing the response as lacking candor. Both the pro hac vice attorney who drafted the filing and the local counsel who sponsored the admission were sanctioned. The court treated the failure as a shared supervisory failure.

Even when courts take a more forgiving view under Rule 3.3, they do not relax supervision duties. In *Garner*, the court declined to find a violation of Rule 3.3 because the attorney did not knowingly submit false authority. But the court still imposed sanctions for the failure to supervise a law clerk who had used AI. Likewise, in *Johnson*, multiple attorneys reviewed a filing containing hallucinated citations. Even though the court declined to extend Rule 3.3, it imposed significant sanctions on all attorneys who signed the pleading, including public reprimands.

Formal Opinion 512 emphasizes that lawyers must review AI-generated work, but it does not fully capture how supervision operates when the "assistant" is not human. The traditional supervision model assumes a process that can be readily reviewed through human involvement. Supervising AI requires more than reviewing the final document and more than telling lawyers to "be careful." Effective supervision requires written policies, training, and verification protocols, all of which must be in place before a problem arises.

At a minimum, firms should adopt a written policy governing the use of AI. That policy should identify which tools are approved for use, explain how those tools have been vetted for security and accuracy, and define what categories of client information may be entered into them. It should require verification of all outputs, prohibit reliance on unverified citations, and state unambiguously that responsibility for the final work product always rests with the lawyer.

Training is equally important, and it cannot be a one-time presentation. At a minimum, lawyers and staff must understand how to structure effective prompts, recognize where confidentiality risks arise, and know what verification is required before relying on any output. More advanced training should address higher-risk uses—such



as litigation strategy, document analysis, legal research, and handling of sensitive data—and may require separate tracks for general users and those working with more sophisticated tools. Lawyers should be trained to treat AI output the way they would treat a draft from a junior associate: verifying every citation and quotation against the source, testing every conclusion before relying on it, and stress-testing assertions in higher-risk work.

Communication (Rule 1.4): Disclosure and Client Expectations

Formal Opinion 512 also addresses communication with clients under Model Rule 1.4 and the reasonableness of fees under Model Rule 1.5. Though these issues attract less attention than competence, confidentiality, or candor, they present some of the most practical issues faced in day-to-day practice.

The Opinion does not impose a blanket requirement that lawyers disclose their use of AI in every matter. Instead, it adopts a contextual approach. According to the Opinion, disclosure may be required where the use of AI is material to the representation, affects the client's risks, or influences how the work is performed. In routine situations, disclosure may not be necessary. In other situations, particularly where sensitive information is involved or where AI plays a substantial role in analysis or strategy, some level of communication may be required.

That sounds straightforward, but in practice, it is not. The line between "routine use" and "material use" is often unclear and will not always be drawn in the same way by the lawyer and the client. A lawyer may view AI as just another tool, no different from Westlaw or Microsoft Word, a perception that is likely to become more typical. A client, however, may view it as a fundamentally different way of performing legal work—one that raises questions about confidentiality, quality, and cost. The line between routine use and material use also is blurring quickly as AI becomes more pervasive.

Formal Opinion 512 leaves the judgment call to the lawyer, which creates risk. The most effective way to manage that risk is not to guess where the line is. The engagement letter is one of the best ways to address the issue. Rather than treating AI disclosure as a case-by-case decision, lawyers should establish a baseline expectation at the beginning of the representation. A clear engagement letter provision can explain how AI may be used in the matter, what safeguards are in place to protect client information, and what effect that use may have on the work and the cost. It can also address the client's right to limit or prohibit AI use, which may require a warning that doing so may result in higher fees when manual labor is required instead of AI.

Disclosing AI use in an engagement letter can accomplish several things at once. It helps to satisfy Rule 1.4 by ensuring clients understand how their matter will be handled, and it avoids the difficult conversation that arises when a client later learns that AI was used without their knowledge. Formal Opinion 512 cautions that boilerplate disclosures alone may not establish informed consent in every case, and that is true. But a well-drafted engagement letter provision sets a clear default understanding at the outset of the representation, which can be supplemented as needed when a specific matter requires more.

Fees (Rule 1.5): Billing, Efficiency, and Value

Fees present a practical challenge under Rule 1.5, one that sits at the intersection of AI efficiency and the traditional hourly billing model. AI can significantly reduce the time required to complete certain tasks, but the value of that work to the client may not have changed. Rule 1.5 requires that fees remain reasonable regardless of how the work is performed. Lawyers cannot bill for time they did not spend, and they cannot treat AI as a hidden profit center by charging "phantom hours" for work that took a fraction of the time it would have taken under traditional methods. At the same time, Formal Opinion 512 recognizes that lawyers may charge for the actual time spent using AI tools and raises the further question of whether flat-fee arrangements will remain reasonable as AI takes over a greater share of the underlying work.

That tension is not fully resolved by the Opinion, and early caselaw suggests it will require further attention. In *White v. Cnty. of Suffolk*, 2024 WL 5221208 (E.D.N.Y. Dec. 26, 2024)—the first reported decision to cite Formal Opinion 512—the court questioned whether the fees at issue might have been more reasonable had the lawyer used AI. The implication is significant: Courts may begin to expect AI-assisted efficiency to be reflected in what lawyers charge. How the profession reconciles hourly billing practices with AI-driven efficiency gains remains an open question, but it is one that will shape the future of legal billing.

Complicating matters further, AI does not always reduce the time a lawyer spends on a task. In many cases, a lawyer may spend the same number of hours but produce a more thorough and comprehensive work product than would have been possible without AI assistance. A lawyer drafting a commercial purchase agreement, for example, might use AI to develop additional representations, closing contingencies, termination rights, or indemnification carve-outs that would have been easy to overlook in a manual drafting process. This can result in a markedly stronger agreement without any reduction in billable time. The same is true in estate planning: A trust agreement drafted with AI assistance might better anticipate distribution contingencies, address potential tax consequences, enhance trustee replacement options, or include more tailored spendthrift provisions than a precedent-based draft alone would have produced. In these situations, the client receives a better document, the lawyer bills the same number of hours, and the fee is almost certainly reasonable. But the dynamic is still worth acknowledging, especially as some clients may begin to expect less time to be spent on tasks. As AI tools become more sophisticated, questions about the relationship among time, quality, and value are likely to become more common in both client conversations and the courts.

Conclusion

ABA Formal Opinion 512 has the fundamentals right: Existing ethical duties govern AI, and the lawyer remains responsible for the work it produces. But the Opinion was written before courts had engaged with these questions in any serious way. The caselaw that has since developed does not contradict the Opinion so much as it exposes what the Opinion could not yet see.

The most consequential gap is Rule 3.3. Candor to the tribunal has become the centerpiece of AI-related sanctions, yet courts are applying it in fundamentally different ways. Whether submitting a hallucinated citation is a per se violation or requires knowing conduct depends entirely on the jurisdiction. That inconsistency means that two lawyers who make the same mistake can face materially different consequences depending on where they practice.

Supervision is no less significant than candor in the emerging caselaw. The cases decided so far make clear that sanctions do not stop with the attorney who used the AI tool, and the sanctions often extend to those in a supervisory position. In *Wadsworth* and *Versant Funding*, supervising attorneys were sanctioned not because they made the error, but because they failed to establish and enforce the processes that would have caught it. The supervision framework that sufficed for human assistants is not adequate for AI. Reviewing the final work product is not enough; supervising attorneys must also oversee the process that produced it. Firms that have not yet established written policies, training programs, and mandatory verification protocols are not simply behind the curve; they are exposed.

The law in this area is still forming, and lawyers who wait for it to settle before adjusting their practices are making a bet that may not pay off. The ethical framework is not going to change in any fundamental way. The duties of competence, candor, confidentiality, and supervision have always required lawyers to understand their tools, verify their work, supervise their subordinates, and take responsibility for the result. AI makes those duties harder to satisfy but not easier to excuse. Lawyers who use AI carefully—verifying outputs, supervising its use, protecting client information, and communicating honestly with clients—will find it a genuinely powerful tool. The technology will continue to improve, and the profession will continue to adapt. But the judgment, verification, and responsibility have always belonged to the lawyer. Nothing about AI changes that. For instance, this article itself was drafted with AI assistance, helping to brainstorm, develop ideas, sharpen prose, and edit. Every source was verified, every output reviewed, and responsibility for the final product remained with the author throughout—competence, supervision, and candor, applied in practice. This article is not just an argument for the ethical use of AI in law; it is also intended as an example of that use. ■