PLI CURRENT

Vol. 10 (2026)

THE TWO LETTERS THAT MATTER MOST IN THE FUTURE OF LEGAL PRACTICE

And They're Not What You Think

Automation Bias, ABA Formal Opinion 512, and the Emerging Ethical Framework for AI-Assisted Legal Work

Erin Corken

Exterro

This two-part series discusses the ethical framework, challenges, and professional obligations surrounding the use of artificial intelligence (AI) in legal practice. The article analyzes automation bias, ABA Formal Opinion 512, evolving sanctions for AI misuse, and provides practical guidance for legal professionals to responsibly integrate AI tools while maintaining ethical standards and competence.

Everyone in the legal profession is talking about two letters: AI. But the two letters that matter most in the future of legal practice are not AI. They are NO, the most underused and most essential word in the age of artificial intelligence. This article examines why, through the lens of automation bias, ABA Formal Opinion 512 (July 29, 2024), a rapidly growing body of sanctions case law, state bar ethics opinions, and court rules. As of June 2026, more than thirty states have issued formal AI ethics guidance, courts from the Southern District of Texas to the Ninth Circuit have developed local rules and issued landmark decisions, and the sanctions for AI-related errors have grown from modest fines to professional suspensions and disqualifications.

Automation bias, the cognitive tendency at the root of nearly every ethical failure in AI-assisted legal practice, runs through all of it. This article provides a complete analysis of Opinion 512, a survey of the evolving landscape since its issuance, practical guidance for practitioners, and a framework for understanding why the most important professional skill in the age of AI is not knowing how to use the technology. It is knowing when to say no to what it tells you.

TABLE OF CONTENTS

I. INTRODUCTION: LIGHT AND SHADE

Everyone in the legal profession is talking about two letters: AI. But the two letters that matter most in the future of legal practice are not AI. They are NO.

Daniela Amodei, co-founder of Anthropic, has described the AI moment as one of "light and shade"[1] and a single day in the life of a legal technology professional and educator in 2026 illustrates exactly what she means.

On the day this article was first drafted, the author used artificial intelligence to produce an initial draft of the very document you are reading, drawing on her own decades of expertise in e-discovery, privacy compliance, and legal education to direct, shape, and verify every element of the output. The light is that a powerful tool amplified her capability. The shade is that the tool she used to draft an article about AI hallucinations is itself capable of hallucinating, which is precisely why she verified every element of it. On that same day, she watched a presentation on AI-powered forensic investigation tools capable of collecting complete records of everything a person does on a computer, every file, every document, every communication, in minutes rather than the days such investigations previously required. The light is that investigations that once took days now take minutes, putting justice within reach faster than ever before. The shade is that the same capability, in different hands or without oversight, can strip privacy bare in the same amount of time. That same evening, she received a legal fee proposal that appeared to roughly double existing rates, and when she asked directly whether the presentation justifying the increase had been prepared with AI assistance, she received what she can only describe as a defensive answer. That is light and shade together: the light is AI potentially creating genuine efficiency and value in legal work; the shade is that practitioners who do not know their ethical obligations regarding whether and how they are permitted to use it may feel the need to deny using it at all. She learned that law students are now being taught to use AI as a standard professional tool in legal writing courses while many of the federal courts where those same students will one day work and practice prohibit entirely its use by the court's staff. The light is that AI makes the work more accessible and the learning curve less daunting. The shade is the uncertainty about whether students are actually learning the real law and not something that was hallucinated, and the institutional contradiction of being trained to use a tool their future employers may prohibit. The generation being trained to use AI and the institutions they will serve are operating

[1] *The Co-Founders of Claude AI Tell Oprah About the Impact Artificial Intelligence Has on Your Life*, The Oprah Podcast (May 19, 2026), https://www.youtube.com/watch?v=w5dJqHilu5s (remarks of Daniela Amodei).

under opposite rules simultaneously. Light and shade, in the same generation, in the same moment.

This is the light and shade Daniela Amodei described. Not a promise. Not a threat. Both. Always both. The professional obligation and the argument of this article is to hold both clearly at once.

The concept has deep roots in psychology. Carl Jung wrote about the capacity to hold the tension of opposites: the ability to keep two opposing truths present at once without collapsing into either, which Jungian psychology considers a mark of psychological maturity.[2] The AI moment demands exactly this capacity. The same technology that makes a lawyer more capable of serving clients also makes that lawyer more vulnerable to professional sanction if that capability is not exercised with judgment. The same tool that democratizes access to legal knowledge also threatens to erode the expertise required to evaluate it. The same efficiency that saves time also creates the pressure to skip the verification that prevents error. These are not contradictions to be resolved. They are tensions to be held. The practitioner's task is not to decide whether AI is good or dangerous, but to hold both truths at once and to choose well, moment by moment.

This is also the framework through which the author approaches her own use of AI in this article. This article was drafted with Claude's assistance. The author directed the work, verified every claim, supplied every personal story, and takes full responsibility for every word. That is light. She also acknowledges that the tool she used to help draft an article about AI hallucinations is itself capable of hallucinating. She checked her own work precisely because she knows the tool cannot be trusted to do it for her. That is shade. Both are true. Both are important. The disclosure in the Appendix is not a confession. It is a demonstration of exactly the practice this article advocates.

This article argues that the defining professional skill of the AI era is not the ability to use artificial intelligence effectively. It is the judgment to know when not to trust what artificial intelligence tells you and the professional courage to say NO.

---

[2] Margaret Mikkelborg, *Holding the Tension of the Opposites*, Psychotherapy Insights (Jan. 17, 2017), https://www.psychotherapyinsights.ca/self-awareness-blog/holding-the-tension-of-the-opposites/.

One vehicle for understanding this framework is ABA Formal Opinion 512, issued July 29, 2024, the American Bar Association's first formal guidance on the ethical obligations of lawyers using generative AI tools.[3] Opinion 512 is, at its core, a document about professional judgment. It does not prohibit lawyers from using AI. It insists that lawyers bring the same ethical obligations to AI-assisted work that they bring to all legal work: competence, confidentiality, communication, bringing only meritorious claims and candor to the tribunal, supervision, and fair dealing on fees.[4]

But Opinion 512 is no longer alone and the stakes of getting this wrong are documented and escalating. In the roughly two years since its issuance, thirty-one states and the District of Columbia have issued formal ethics guidance on AI in legal practice[5], courts have developed local rules and standing orders, and the sanctions landscape has escalated from modest fines to five-figure penalties, professional suspensions, attorney disqualifications and bar referrals. Over 1,400 court decisions worldwide have addressed AI-generated errors in legal filings, with ninety percent of those decisions written in 2025 alone.[6] Courts have made clear that ignorance of AI's capacity to fabricate citations is no excuse. This article addresses all of it.

Underlying this sanctions landscape is a psychological phenomenon that the legal profession has been slow to name and even slower to address: automation bias, the documented human tendency to over-rely on automated systems and accept their outputs without sufficient critical evaluation.[7] This is the reason that intelligent, experienced lawyers are submitting fabricated citations to courts. It is not carelessness in the ordinary sense. It is the predictable result of human cognitive architecture

---

[3] ABA Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024) https://www.americanbar.org/news/abanews/aba-news-archives/2024/07/aba-issues-first-ethics-guidance-ai-tools/ (Generative Artificial Intelligence Tools).

[4] *Id.*

[5] *State Bar AI Guidance*, Legal AI Governance, legalaigovernance.com/tracker/states/ (last visited July 7, 2026).

[6] Damien Charlotin, *AI Hallucination Tracker*, https://www.damiencharlotin.com/hallucinations/ (last visited July 14, 2026).

[7] E.R. Jones, *Automation Bias*, The Decision Lab (2025), https://thedecisionlab.com/biases/automation-bias.

encountering a technology specifically designed to produce outputs that look authoritative and correct.

Something else has changed since Opinion 512 was written, and it runs deeper than any advance in the technology itself. The way human beings interact with their technology has fundamentally shifted, and this author argues it is one of the more transformative changes humankind has quietly lived through. For years, searching the internet meant entering a query and receiving a list of websites to open, read, and evaluate. The verb for it became a household word: to Google something. That is no longer what happens. Today the same search returns a synthesized answer, the information already gathered, read, and summarized, as though the work of clicking through and reading had been done for us. And it does not stop there. We can ask a follow-up question, and the technology answers back. Whether or not we select a tab labeled AI, we are now conversing with our tools, and we have accepted this as unremarkable.

We did not pause. We did not stop to ask what was happening. A few years ago, at a conference session on AI, an audience member asked the speaker how one goes about buying AI and putting it to use. The speaker replied that you do not buy AI; it arrives as a software update to the technology you already have. That is no longer entirely true, as there are now many ways to purchase and deploy these tools deliberately. But as a description of how AI entered most people's daily lives, it was exactly right. It simply appeared, and we absorbed it without reflection, without challenge, without verification, and without ever saying NO, because we like it. It is easy, and it is useful, and it asks nothing of us.

Consider how strange this would seem in any other context. If the author needed to research tablecloths, she would not pose questions to a tablecloth and expect it to answer, and if it did, no one would treat that as ordinary. Yet a version of exactly that now happens every day, at scale, and it has not given us pause. That absence of pause is itself the point. This article is, in some measure, that pause.

It proceeds in eight parts. Part II provides a plain-language overview of the technology: generative AI, agentic AI, where we currently stand in AI's development, and two types of hallucinations. Part III examines automation bias and the psychological tendencies at the heart of the problem. Part IV analyzes ABA Formal Opinion 512. Part V surveys the sanctions landscape and the accelerating body of cases involving AI errors. Part VI turns to the humanity at the center of it all. Part VII offers practical guidance including checking rules, creating policies, and verifying every output. Part VIII concludes.

II. THE TECHNOLOGY: SOME CLARIFICATION

Because we are now conversing with these tools rather than merely querying them, understanding what they actually are matters more than ever.[8] The shift described in the Introduction, from receiving a list of sources to receiving a synthesized answer that talks back, means most people are already using AI fluently without understanding how it works or where it fails. This Part offers a plain-language account of the technology: what generative AI is, what agentic AI is, where we actually stand in AI's development, and two kinds of hallucination errors these tools produce.

A. What Generative AI Is and How It Works

Generative artificial intelligence refers to AI systems that generate text, images, code, or other content in response to natural language prompts. Unlike traditional software, which follows explicit programmed instructions, generative AI tools are trained on vast datasets of human-generated text and learn to predict statistically likely continuations of a given input. The large language models (LLMs) that power tools like ChatGPT, Claude, Gemini, and Microsoft Copilot do not "know" facts in the way that a database knows facts. They generate text that is statistically probable given their training data. This distinction is crucial for understanding both the power and the danger of these tools.

Understanding what generative AI is and is not is the foundation. But the AI landscape does not hold still. In the two years since ABA Formal Opinion 512 was issued in July 2024, the technology has continued to evolve in ways the Opinion did not anticipate and that every practitioner must now understand.[9]

B. Agentic AI: The Next Frontier

ABA Formal Opinion 512 focused on generative AI as it existed in mid-2024, tools that respond to prompts and generate text. But the AI landscape has evolved. Agentic AI refers to autonomous AI systems designed to achieve complex goals. It is more than a passive system that simply generates content, predicts text or answers

[8] AI capabilities are evolving rapidly. Descriptions of specific tools and their limitations in this article reflect the state of the technology as of the date of publication.

[9] ABA Formal Op. 512 n.4. The Opinion acknowledged that additional issues may surface as the technology develops. The developments this article describes are among them.

questions; it is designed to take independent action across multiple steps, making decisions and taking actions along the way without continuous human direction.

An agentic tool might be given an objective such as researching an issue, drafting a filing, or responding to an inquiry and it autonomously executes the necessary actions to achieve it. It does things such as navigating browsers, retrieving documents, writing code, and sending communications, independently deciding which steps to take and when. In advanced implementations, agentic AI operates as a coordinated team: a lead orchestrator evaluates the goal, plans the workflow, and assigns tasks to specialized sub-agents, dynamically pivoting as the work develops, to synthesize and complete the assignment.

In a legal context, this means an agentic AI tool could conduct multi-source legal research, draft a brief based on what it found, and flag it for attorney review, all without the attorney having directed each individual step. The efficiency gains are significant. So are the risks. The tool is making judgment calls along the way, and those judgment calls may not be visible in the final output. An error introduced in an intermediate step, a misread source, a missed authority, a flawed inference can travel invisibly through the process and surface in a filing the attorney never saw being built. This is why transparency in how agentic tools operate, and the knowledge of how to use them responsibly, are not optional features. They are the difference between a powerful tool and an uncontrolled one. Understanding that difference begins with understanding where these systems actually sit on the spectrum of AI development.

C. The AI Spectrum: ANI, Broad ANI, AGI, and ASI - Where We Actually Are

One of the most persistent and consequential misconceptions about artificial intelligence is the assumption that current AI systems are, or are close to being, the kind of sentient, all-knowing machines depicted in science fiction. People often ask the author whether they should worry about AI taking over. Her answer: we are not living in the Matrix, yet, at least not that anyone has confirmed. Understanding where we actually are on the AI spectrum is essential to understanding both the power and the limitations of the tools this article addresses.

Artificial Narrow Intelligence (ANI) is what most AI has been so far, systems designed for one specific task and one task only. A spam filter. A chess engine. A voice recognition system. Broadly Capable Narrow Intelligence sometimes called Broad ANI is where we actually seem to be today with tools like Claude, ChatGPT, Gemini,

and Copilot. These systems are still technically narrow.[10] They are not conscious in any way science can currently verify. But they are capable across a surprisingly wide range of language and pattern recognition tasks. The breadth of capability is real and impressive.

Further along the spectrum sits Artificial General Intelligence (AGI), AI that can reason, learn, and apply intelligence flexibly across any domain the way a human can. Whether any current system has reached that threshold is a question researchers have not resolved. Artificial Super Intelligence (ASI), the scenario that animates science fiction and philosophical debate, does not exist in any form science has documented. What lawyers need to understand is more practical: the automation bias problem is not a superintelligence problem. It is a Broad ANI problem. And the most consequential way that problem shows up in legal practice has a name.

D. Hallucinations: Two Types of AI Error

The legal profession has adopted the term "hallucination" from AI research and development to describe the phenomenon of AI generating fabricated information presented with apparent confidence. In AI research, the term was coined to describe this specific failure mode.[11] The legal profession has widely adopted it as these tools have become embedded in legal practice and recently the profession has made some important distinctions about types of hallucinations that we will look at below. But to understand those distinctions better we first must look at why AI makes things up.

---

[10] Harish Tayyar Madabushi, Melissa Torgbi & Claire Bonial, *Neither Stochastic Parroting nor AGI: LLMs Solve Tasks through Context-Directed Extrapolation from Training Data Priors*, arXiv (2025), https://arxiv.org/html/2505.23323v1. These systems may still be technically narrow, though researchers disagree about where exactly they fall on the spectrum and the nature of whatever processing they perform remains genuinely contested.

[11] Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. Empirical Legal Stud. 216, 221 (2025), https://law.stanford.edu/publications/hallucination-free-assessing-the-reliability-of-leading-ai-legal-research-tools/. Fabrications are instances in which the generative AI tool provides cases or quotations that do not exist at all. *Id.* at 221, 230. Inaccuracies are more subtle. *Id.* at 222. The generative AI tool might cite to real authorities but provide an answer that is legally or factually inaccurate or not supported by the citation.

The answer lies in what these tools are actually doing when they generate text. A large language model, at its core, does not look up facts. It predicts the next most statistically likely word, over and over, based on patterns it learned from enormous amounts of text.[12] Some legal research tools add a retrieval step that feeds the model real source material before it answers, an approach discussed below, but the engine underneath is still predicting text, which is why even those tools can err. When a model produces a case citation, it is not retrieving a real case. It is generating a string of text that looks like a citation, because it has seen millions of real citations and has learned their shape: a case name, a volume number, a reporter, a page, a court, a year.

That is the heart of the problem. The model is extraordinarily good at reproducing the form of a citation and entirely indifferent to whether anything real sits behind it. A fabricated case is not a glitch or a malfunction. It is the system doing exactly what it was built to do, which is generate plausible text. The plausibility is precisely what makes it dangerous. A citation that looks wrong would be caught immediately. A citation that looks perfect, with a real-sounding name and a properly formatted reporter cite, sails through unless someone checks whether the case actually exists.

This is why the profession has begun paying closer attention not just to whether AI hallucinates, but to the different forms its errors take. Until recently, "hallucination" was used primarily to describe fabrications, cases and citations that do not exist. But in June 2026, the United States Court of Appeals for the Ninth Circuit provided the legal profession with a more precise taxonomy in *Lnu v. Blanche*.[13] The court distinguished two types of AI error that every practitioner must understand.[14] The first type is fabrication: AI invents entirely fictional judicial opinions from scratch, complete with

---

[12] *See 1.1.1: How Language Models Generate Text by Predicting the Next Word*, SOC. SCI. LIBRETEXTS, https://socialsci.libretexts.org/Courses/Merced_College/Demystifying_AI%3A_A_Practical_Introduction_for_Instructors/01%3A_An_Introduction_to_AI/1.01%3A_What_is_a_language_model/1.1.01%3A_How_language_models_generate_text_by_predicting_the_next_word (last visited June 14, 2026); *A Legal Practitioner's Guide to AI & Hallucinations*, NAT'L CTR. FOR STATE CTS., https://www.ncsc.org/resources-courts/legal-practitioners-guide-ai-hallucinations (last visited June 14, 2026).

[13] *Malkeet Lnu v. Blanche*, No. 24-4790, 2026 U.S. App. LEXIS 16174 (9th Cir. June 3, 2026).

[14] *Id.* at *10.

fake case names, fake reporter volumes, and fake page numbers.[15] These cases do not exist and never existed. The second type is inaccuracies: AI "might cite to real authorities but provide an answer that is legally or factually inaccurate or not supported by the citation."[16]

Both types share the same root cause, and two features of how these tools work make both kinds of error easy to miss. First, AI Sycophancy. These tools are designed to be helpful and responsive, so when you ask for cases supporting your position, they supply them, whether or not such cases exist. It is simply giving you what your prompt asked for in the form you asked for it. Second, nothing in the process checks the output against reality before it reaches you. Unless the tool is connected to a verified database as discussed below, there is no ground truth, no moment where the system confirms that what it just wrote is real. That confirmation is your job. It always has been.

Now, even connecting the tool to a verified database does not solve the problem, as the legal profession has learned. Stanford University's RegLab, in a study cited directly in both Lnu and ABA Formal Opinion 512, found that leading AI legal research tools generate hallucinations, including both fabrications and inaccuracies, in between seventeen and thirty-three percent of queries.[17] That number is significant, because these were not consumer chatbots. The tools studied included the leading legal research platforms, such as Lexis+ AI and Westlaw's AI-Assisted Research, which are built specifically for lawyers. As noted above, these tools add a retrieval step, an approach called retrieval-augmented generation, or RAG, that connects the model to a curated database of real legal authorities so it can draw from verified sources rather than from patterns alone. RAG is an AI architecture that allows a language model to search an external database for information before generating a response, operating much like an "open-book" exam. Enterprise or Domain-Specific RAG limits this search to a closed, highly vetted, and specialized repository of data, such as a provider's proprietary database of verified case law, statutes, and regulations. In contrast, Real-

---

[15] *Id.* at *14. In *Lnu v. Blanche*, the brief cited nonexistent decisions titled "Eduardo v. Garland, 28 F.4th 742 (9th Cir. 2022)" and "Lay v. Holder, 729 F.3d 962 (9th Cir. 2013)."

[16] *Id.*

[17] ABA Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024) (citing Varun Magesh et al., *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, 22 J. Empirical Legal Stud. 216 (2025), https://law.stanford.edu/publications/hallucination-free-assessing-the-reliability-of-leading-ai-legal-research-tools/).

Time or Web-Search RAG pulls live, unvetted information dynamically from across the open internet to address real-world events or breaking news. RAG meaningfully lowers the hallucination rate. But the engine underneath is still the same predictive model, which is why it does not eliminate the problem. Even these purpose-built, database-connected legal tools still produced fabrications and inaccuracies in a meaningful share of queries. No tool, however sophisticated or legal-specific, can be trusted to catch its own errors. The question is whether the human reviewing the output will catch them instead and the answer depends on a psychological phenomenon the legal profession has been slow to name.

III. AUTOMATION BIAS: WHERE ETHICAL MISTAKES BEGIN

A. Defining Automation Bias

Automation bias is a well-documented psychological phenomenon defined as the tendency to over-rely on automated systems and accept their outputs without sufficient critical evaluation even when the automated output contains errors that a reasonably attentive human reviewer would detect. The phenomenon was first systematically described in the aviation context and has since been documented in medical, financial, and legal contexts.

The core mechanism of automation bias is not ignorance or carelessness. It is a predictable response to cognitive load. When a person is managing multiple demands on their attention as lawyers routinely are automated tools that provide seemingly complete, authoritative outputs reduce the cognitive burden of the task at hand. This clarifies why professionals would engage in behaviors they might otherwise avoid, such as failing to verify output accuracy. It is the predictable outcome of human cognitive architecture encountering a tool specifically designed to reduce cognitive effort.

B. The Confirmation Bias Amplification Effect

Confirmation bias, the documented human tendency to search for and interpret information in ways that confirm existing beliefs, compounds automation bias.[18] Lawyers do not approach legal research as neutral observers. They approach it with a goal: to support a position. That goal shapes what they look for and what they are

[18] Jones, *supra* note 7.

inclined to accept. When research appears to confirm the position, a lawyer already holds, the lawyer is less likely to scrutinize it. Automation bias creates over trust in the tool; confirmation bias adds the motive to believe what the tool returns.

C. Why Legal Professionals Are Particularly Vulnerable

It is tempting to hand the harder things in life over to AI, the way humanity does in *WALL-E*, the 2008 Disney-Pixar film set in a distant future. Aboard the starship Axiom, generations of humans have grown physically frail and wholly dependent on the machines that do everything for them, drifting through life in hovering chairs, never walking, never choosing, never struggling with anything, just floating around sipping smoothies. The film is a warning. It shows how technology, left unchecked and unquestioned, can quietly strip away human agency, health, and connection until people forget they ever had them. It is a vivid picture of the erosion this article takes up later: the gradual loss not only of knowledge, but of capability itself.

The temptation to look for the easy button does not arise in a vacuum. It is amplified by cultural pressures that make deference feel not only natural, but virtuous. Dale Carnegie, in *How to Win Friends and Influence People*, cautions against telling someone they are wrong.[19] Our broader culture teaches us to avoid saying no, lest we appear difficult, obstructionist, or unkind. People-pleasing is a recognized pattern that affects professionals across every field. A former colleague in sales once shared that he had made it a personal rule never to use the word "no," not once, not ever. The pressure to accommodate, to smooth things over, to accept rather than challenge, is pervasive. And when the thing being deferred to is a machine that speaks with confident authority, that pressure becomes even harder to resist.

The temptation to surrender control is not unique to lawyers, but lawyers are arguably particularly vulnerable to it and for reasons that are deeply embedded in how the profession trains and shapes them from the very beginning.

In law school, students are taught through the Socratic method. A professor calls on a student without warning, poses a question, and then regardless of the answer poses another question, probing, pressing, exposing every gap and assumption in the student's reasoning. The purpose, in theory, is to develop rigorous analytical thinking, to teach students to hold an argument up to the light and examine it from every angle.

[19] Dale Carnegie, *How to Win Friends and Influence People* (1936).

It is an intellectually demanding and often humbling process, and it has genuine value. That is the light. But it also carries shade.

The Socratic method trains lawyers to fear being wrong in public. It teaches, at a formative stage, that intellectual exposure is dangerous, that saying the wrong thing in front of an authority figure carries real consequences. And those consequences turn out to be lasting. First-year law school grades, typically assigned at the end of the year after a single set of high-stakes examinations, are enormously determinative of a lawyer's future earning potential. The most prestigious law firms recruit almost exclusively from the top of the class, at precisely that moment, before a student has had the opportunity to demonstrate growth, resilience, or wisdom accumulated over time. One set of exams, at one moment in time, can define a career trajectory for decades.

The result is a profession trained simultaneously to demand intellectual rigor and to dread the appearance of imperfection. If you have worked alongside attorneys without being one yourself, you may have noticed a certain tension, a reluctance to admit uncertainty, a tendency to project confidence even when the ground is uncertain. Now you know some of the reasons why.

This dynamic makes lawyers unusually susceptible to AI's most seductive quality: it never hesitates. It never says, "I'm not sure." It produces polished, confident text at speed, with the reassuring appearance of authority. For a professional conditioned to treat hesitation as weakness and imperfection as professionally costly, that quality is intoxicating and dangerous.

To this vulnerability add the structural pressures of legal practice itself: the billing models that reward speed, the client expectations that demand responsiveness, the court deadlines that compress careful thought into shrinking windows of time, and the sheer cognitive load of managing complex matters simultaneously. In that environment, efficiency is constantly rewarded and thoroughness is sometimes penalized. It is no wonder that lawyers occasionally cut corners they should not, including the corner of verification. And cutting that corner is not new. Courts were reckoning with the consequences of counsel's carelessness, inattention, or overconfidence long before AI entered the equation.

The author's experience as a federal judicial law clerk illustrates the point. Clerking for the Chief Judge, her responsibilities included reading every pleading submitted by counsel and checking each legal argument against the authorities cited. One afternoon, working through the briefs for a newly filed case, she found herself rereading the same page repeatedly, unable to make sense of it. Convinced after a long day that the problem was with her own comprehension, she walked down the hall to a more

experienced colleague and handed the page over without explanation. The colleague read it quickly, handed it back, and said matter-of-factly: it's gibberish. A page of words that formed no coherent sentences, random, disconnected, entirely meaningless, and it had been filed in the United States Federal District Court by a licensed attorney.

Attorneys, in other words, have always been capable of filing documents without reading them carefully, or at all. AI did not create the failure to verify. What it has done is make that failure faster, more plausible-looking, and far more consequential. It tempts us toward exactly the *WALL-E* future, letting the machine handle the hard things at precisely the moment we can least afford to. As the technology section discussed, we are not quite there yet. We are working alongside tools that make errors as readily as we do, and sometimes more. To hand such a tool the things that matter most, without keeping a hand on the wheel ourselves, is not efficiency. It is abdication.

The profession has begun to respond. The first and most authoritative response came from the American Bar Association.

IV. ABA FORMAL OPINION 512: THE NATIONAL FRAMEWORK

A. Overview and Context

ABA Formal Opinion 512, issued July 29, 2024, is the American Bar Association's first formal opinion specifically addressing the ethical obligations of lawyers who use generative AI tools. Its central message is one this article has already begun to make: AI did not create new duties. It is testing old ones. The Opinion does not invent a single new rule. It establishes, with specificity and authority, that the same professional obligations lawyers have always borne apply with full force to AI-assisted work, and that the duties to be competent, to protect client confidences, to communicate with clients, to be candid with the court, to supervise those one is responsible for, and to charge reasonable fees do not soften because a machine drafted the document, handled the client's information, or did work the lawyer would otherwise have done.

The Opinion addresses six ethical obligations spread across several model rules: Competence (Rule 1.1), Confidentiality (Rules 1.6, 1.9(c), and 1.18(b)), Communication (Rule 1.4), Candor Toward the Tribunal (Rules 3.1, 3.3, and 8.4(c)), Supervisory Responsibilities (Rules 5.1 and 5.3), and Fees (Rule 1.5).[20] This article

[20] ABA Formal Op. 512.

focuses most closely on the obligations that turn on verification: competence, candor, and supervision, because those are where automation bias does its damage. But all six are presented here so this overview of Opinion 512 is complete, and because each is, at bottom, a place where the lawyer must say NO: no to unverified output, no to the unsafe tool, no to invisible use, no to the unchecked filing, no to unsupervised delegation, no to dishonest billing.

Six obligations, six refusals, one word. And while ABA formal opinions are not binding on state bars or courts, they establish the standard of care against which lawyer conduct is measured. Opinion 512 was the first comprehensive statement of that standard for AI. The more specific bar opinions, court rules, and guidance discussed in Part VII came later, and they vary in their particulars, but they all follow the same basic foundation that this Opinion lays out.

B. Competence — Model Rule 1.1

Model Rule 1.1 requires lawyers to provide competent representation, including the legal knowledge, skill, thoroughness, and preparation reasonably necessary for the representation.[21] Comment 8 extends this obligation to technology, requiring lawyers to keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology.[22]

ABA Formal Opinion 512 addresses competence in two dimensions: understanding the capacity of generative AI tools, and more importantly for the automation bias analysis understanding their limitations.[23] A lawyer who uses an AI tool without understanding that it can generate fabricated citations is not competent in the relevant sense. Practitioners need to know that even legal-specific AI tools hallucinate.

---

[21] Model Rules of Prof'l Conduct r. 1.1 (Am. Bar Ass'n 2025).

https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_1_competence/

[22] Model Rules of Prof'l Conduct r. 1.1 cmt. 8 (Am. Bar Ass'n 2023).

https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_1_competence/comment_on_rule_1_1/

[23] ABA Formal Op. 512 at 2–6.

Understood this way, competence in the age of AI is not only the ability to use these tools, it is the judgment to know when to say NO to what they produce. The duty of competence and the duty to verify are the same duty: a lawyer who cannot recognize when an AI tool is likely wrong, and who does not check, has potentially not met the standard Rule 1.1 requires, however fluent the output appears.

C. Confidentiality — Model Rules 1.6, 1.9(c), and 1.18(b)

Model Rule 1.6 requires lawyers to protect confidential client information. Rules 1.9(c) and 1.18(b) extend these obligations to former clients and prospective clients respectively.[24] Together these rules create a comprehensive confidentiality framework that applies to AI tool use.

There are different types of AI tools that a practitioner might use: consumer and enterprise. The distinction between consumer and enterprise AI tools is not merely a technical detail. It has direct legal consequences that can affect a client's rights. The lawyer who does not understand this distinction cannot give competent advice about AI use to clients.

Consumer versions of AI tools typically retain user inputs and use them to improve the underlying model. This means anything a lawyer enters, such as client names, case facts, draft arguments, or settlement positions, leaves the lawyer's control and is absorbed into a system the lawyer neither owns nor governs. The information may be stored indefinitely, reviewed by the provider's personnel, used to train future versions of the model, and exposed through the model's later outputs. Once confidential information is entered into such a tool, the lawyer can no longer assure the client that it remains confidential, and that loss of control is itself the harm Rule 1.6 exists to prevent. The decision reinforces the critical distinction between consumer and enterprise AI tools for confidentiality purposes: using a public AI platform may not

[24] Model Rules of Prof'l Conduct r. 1.6; r. 1.9(c); r. 1.18(b) (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_6_confidentiality_of_information/.https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_9_duties_of_former_clients/. Rule 1.18: Duties to Prospective Client

only violate Rule 1.6, it may waive privilege over the resulting materials entirely.[25] So the first NO regarding confidentiality is simple: NO to putting client confidences into a consumer tool. Absent the client's informed consent, NO.

Enterprise versions typically provide contractual protections including things such as commitments not to use inputs for model training, data deletion provisions, and security certifications. They also typically segregate data by client and matter and do not commingle data. These protections address many of the concerns the Opinion raises about cross-contamination of information between clients and matters. In practical terms, an enterprise tool performs work on one matter and does not carry that information into another, much as a lawyer's duty of confidentiality walls one client's information off from another. But because these protections vary by vendor, the lawyer must read and understand the contract to confirm what a given tool actually does with client data. So, the second confidentiality NO is conditional: NO to assuming an enterprise tool is safe until the contract has been read and its protections confirmed. The word "enterprise" alone guarantees nothing.

D. Communication — Model Rule 1.4

Model Rule 1.4 addresses a lawyer's duty to communicate with clients.[26] Regarding AI, the Opinion identifies circumstances where disclosure is required: when a client asks how the work was done and AI was used, when a client asks whether AI was used, and when the client "expressly requires disclosure under the terms of the engagement agreement or the client's outside counsel guidelines."[27] Outside those situations, disclosure may or may not be necessary, and each use should be evaluated case by case. The Opinion offers some guidance on where disclosure is more likely to be warranted, such as consulting the client before using AI to evaluate litigation outcomes or assist with jury selection, or where the client expects the attorney to apply

---

[25] *United States v. Heppner*, No. 25-cr-00503-JSR (S.D.N.Y. Feb. 10, 2026) (holding that documents generated using the consumer version of Claude were not protected by attorney-client privilege or the work product doctrine).

[26] Model Rules of Prof'l Conduct r. 1.4 (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_4_communications/.

[27] ABA Formal Op. 512, at 9.

their own particular skill and knowledge. But it stresses that each circumstance must be considered on its own.

This case-by-case standard is workable, but it leaves the lawyer to make a judgment call every time, and a missed call is a communication violation. There is a simpler and more honest course. In the age of AI, saying NO regarding communication means refusing to let the use of these tools stay invisible. Say NO to feeling like you have to hide that you used it. Disclose it proactively. Address AI use in the engagement letter before the question is ever asked. This is the lawyer choosing accountability over the quiet shortcut, and choosing not to be afraid of getting found out. Which brings us to the next rule, candor toward the tribunal.

E. Candor Toward the Tribunal — Rules 3.1, 3.3, and 8.4(c)

Candor toward the tribunal is where the consequences of automation bias become most visible and most severe. Three rules work together here. Rule 3.1 establishes that a lawyer cannot file baseless lawsuits.[28] Rule 3.3 prohibits a lawyer from knowingly making a false statement of law or fact to a tribunal, and requires the lawyer to correct any material false statement previously made.[29] And Rule 8.4(c) prohibits conduct involving dishonesty, fraud, deceit, or misrepresentation[30], which, as the Opinion notes, can reach even an unintentional misstatement to a court.[31] ABA Formal Opinion 512 is unequivocal: attorneys must carefully review AI outputs to ensure that representations made to a tribunal are not false or misleading, and they must correct any such representations previously made.[32]

---

[28] Model Rules of Prof'l Conduct r. 3.1 (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_3_1_meritorious_claims_contentions/.

[29] Model Rules of Prof'l Conduct r. 3.3 (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_3_3_candor_toward_the_tribunal/.

[30] Model Rules of Prof'l Conduct r. 8.4(c) (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_8_4_misconduct/.

[31] ABA Formal Op. 512, at 10.

[32] *Id.*

A couple of things are worth noting here. First, the verification obligation cannot be satisfied by asking AI to verify its own output. AI tools will typically confirm their own fabrications as accurate. Second, many of the cases that discuss candor to the tribunal share a pattern: an attorney used AI, did not proactively disclose it as recommended above, and then denied it when the court asked. It is no surprise, then, that a growing body of local court rules now requires AI disclosure.

So, follow the same instinct that governs disclosure to clients: when a court requires disclosure, or asks you directly, disclose. You can correct an honest mistake. Even brilliant attorneys make mistakes. No one is above error. What you cannot easily recover is the court's patience, which runs understandably thin for the attorney who will not come clean when asked. Take it from a former United States District Court clerk who used to be the one spending extra time dealing with attorneys who did not follow the rules: do not test the court's patience.[33] Say NO to irritating the tribunal.

F. Supervisory Responsibilities — Rules 5.1 and 5.3

Two rules work together here. Rule 5.1 provides that lawyers with managerial authority in a firm must make reasonable efforts to ensure the firm has measures in place giving reasonable assurance that all its lawyers conform to the rules of professional conduct, and that supervising lawyers must make reasonable efforts to ensure that the lawyers they oversee do the same.[34] Rule 5.3 extends those same supervisory duties to nonlawyer assistance, requiring lawyers to make reasonable efforts to ensure that the conduct of nonlawyers they work with is compatible with the lawyer's own professional obligations.[35]

---

[33] This is the same advice I give my law students. I cannot speak for other professors, but I want my students to do well, and giving a student a non-passing grade is a great deal of extra work for me. Believe me, I am trying everything I can to pass all of them. Work with your teachers, and work with the courts, to help them help you.

[34] Model Rules of Prof'l Conduct r. 5.1 (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_5_1_responsibilities_of_a_partner_or_supervisory_lawyer/.

[35] Model Rules of Prof'l Conduct r. 5.3 (Am. Bar Ass'n 2025), https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_5_3_responsibilities_regarding_nonlawyer_assistant/

Opinion 512 applies these duties to AI in the predictable ways. Managerial lawyers must establish clear policies on the firm's permissible use of generative AI.[36] Supervisory lawyers must make reasonable efforts to ensure that the firm's lawyers and nonlawyers comply with their professional obligations when using these tools, and must see that those people are trained, in the basics of the technology, in the capabilities and limitations of the specific tools, in the ethical issues their use raises, and in secure handling of data, privacy, and confidentiality. These duties are tested most by what is sometimes called shadow AI: lawyers and staff using AI tools the firm has not approved and may not even know about. The policy and training requirements exist precisely because of it. A policy no one is told about, and training no one receives, leaves exactly the gap shadow AI slips through, the unapproved tool, used quietly, on client work, with no one supervising the result. Closing that gap is the whole point of the supervisory duty. The Opinion also treats AI providers much as it treats any outside vendor entrusted with client information, instructing lawyers to vet a tool's reliability, security, data-retention practices, and limitations before relying on it.[37]

All of that is sound, and all of it is about supervising people and vetting vendors. But it leaves a gap, and the gap is where this author offers an argument of her own. Rules 5.1 and 5.3 are written to govern how a lawyer supervises other humans and the outside services a firm engages. They were not written with a tool in mind that produces work product the way a junior associate does. And yet that is exactly what agentic AI does. It drafts, it researches, it answers, it produces work that looks like the work of a person. In this author's view, the supervisory principle behind these rules, that a lawyer remains responsible for work done in their name and must review it rather than trust it, applies with full force to what an AI tool produces, even though the Opinion does not frame it that way.

In suggesting that the principle behind Rules 5.1 and 5.3 should reach the work product generated by an AI tool, the author does not suggest the Committee erred. The Committee had it right for its time, and it expressly recognized that the technology was evolving rapidly and that the profession's ethical approach would need

[36] ABA Formal Op. 512, at 10.

[37] ABA Formal Op. 512, at 11.

to evolve with it. Extending the principle as the technology develops is precisely what the Committee anticipated the profession would have to do.[38]

This article suggests that to bridge the gap between generative AI and agentic AI, the approach to ethics should focus not only on verification but also supervision. When AI takes autonomous actions on behalf of a lawyer or client, filing documents, sending communications, conducting research across multiple sources, it produces work product the way a person would, and the work product needs to be supervised accordingly.

The most useful way the author has found to honor that principle is simple: treat the AI like a direct report. You do not hand a new associate's memo to a court without reading it. You do not assume a paralegal's research is complete because it looks thorough. You supervise the work because you are the one who answers for it. An AI tool deserves exactly that posture, and for the same reason. A lawyer who deploys an agentic AI tool cannot simply review the final output; the lawyer must understand what actions the tool took along the way. And, the hallucination problem does not disappear in Agentic AI; it compounds because errors made in intermediate steps may not be visible in the final output. So, a combination of the requirements of Rules 1.1 and 5.1 and 5.3 is required.

The duty here, then, is the supervision version of the word that runs through this entire article: say NO to unsupervised reliance, on the tool and by the people using it. Whoever or whatever produced the work, a human reviews it, and a human answers for it.

G. Fees — Model Rule 1.5

Rule 1.5 prohibits a lawyer from charging an unreasonable fee or unreasonable expenses, and requires the lawyer to communicate their fees to the client.[39] The law firm billing conversation the author described in the Introduction illustrates the practical stakes. A firm that uses AI to compress work that once took hours into minutes, and then bills as if the AI had not been used, has violated Rule 1.5 as clearly

---

[38] ABA Formal Op. 512 n.4, *supra* note 9.

[39] Model Rules of Prof'l Conduct r. 1.5 (Am. Bar Ass'n 2025),https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_5_fees/

as if it had fabricated the time entries. The method of inflation is new. The ethical violation is not.

Opinion 512 applies the rule to generative AI in several specific ways.

First, hourly billing: a lawyer who bills by the hour may bill only for time actually spent. If an AI tool reduces a task that once took three hours to thirty minutes, the lawyer bills for thirty minutes.[40]

Second, flat and contingent fees: if an AI tool lets a lawyer complete the work far faster, it may be unreasonable to charge the same flat fee that assumed the old, slower effort. As the Opinion puts it, a fee charged for which little or no work was performed is an unreasonable fee.[41]

Third, expenses and overhead: AI costs are sorted into two existing categories, and the long-standing rules for each apply. A lawyer may not bill general office overhead, the ordinary cost of running a practice, and may not add a surcharge to a genuine out-of-pocket expense, billing it instead at actual cost. The Opinion directs lawyers to analyze each AI tool's characteristics and uses, because they vary so widely, and to classify the cost accordingly. A tool that functions like the ordinary equipment of a practice is overhead: a grammar checker built into the lawyer's word processing software, for example, may not be separately charged to the client absent advance disclosure, just as office supplies may not. But a third-party AI service billed per use, such as a tool that reviews thousands of contracts for a particular client, is a genuine out-of-pocket expense: the lawyer may bill the client for the actual cost of that use, but, as with a court reporter or travel, may not surcharge it.[42]

Fourth, in-house services: in-house services such as photocopying are not general office overhead. A lawyer may charge the direct cost of the service plus a reasonable allocation of overhead connected with providing it, or another amount only if agreed on in advance. Proprietary in-house AI tools fall in this category. Whatever is charged must not be duplicative of other charges to this or any other client.[43]

Finally, learning how to use AI: a lawyer may not charge for the time spent learning how to use an AI tool when that time is necessitated by the lawyer's own inexperience,

---

[40] ABA Formal Op. 512, at 12-14.

[41] *Id.* at 12.

[42] *Id.* at 13.

[43] *Id.*

because lawyers must maintain competence in the tools they use. The exception is when the client requires the lawyer to use a specific tool the lawyer does not yet know; there, the lawyer may charge for the learning time. A lawyer is not required to bill for the training, and many firms will not, recognizing the value of becoming proficient with a new tool and the opportunity it represents to gain valuable experience. A lawyer who does choose to bill for it should agree with the client on the terms in advance, preferably in a formal agreement.[44]

The fee rules, like the others, come down to a refusal. The efficiency AI creates is real, and so is the temptation to capture it: to bill the old three hours for the new thirty minutes, to fold a personal education into a client's invoice, to pass a tool's cost through without saying so. But the time the tool saved belongs to the client, not the lawyer. Rule 1.5 asks the lawyer to say NO to all of it: NO to charging for time not spent, NO to billing the client for your own learning, NO to undisclosed or duplicative costs. The honest bill reflects the work actually done, by the lawyer and the tool together, and nothing more.

This is part I in a series.

Erin Corken, Esq. CEDS · FIP · AIGP · CIPP/US · CIPP/E · CIPM is an attorney licensed in Kentucky and Arizona, a law professor and scholar, and a Senior Solutions Engineer and Academic Program Developer at Exterro. A former law clerk to the Chief Judge of the United States District Court for the District of Arizona, she has for more than a decade taught E-Discovery and Information Privacy and Cybersecurity at the law school level, along with Business Law at the undergraduate level, and she has presented at bar associations, paralegal organizations, and legal technology conferences nationwide, including PLI. She has published previously in the Northern Kentucky University Chase Law Review, the John Marshall Law Journal, Bloomberg BNA's Digital Discovery & e-Evidence reporter, and other publications. The author thanks Ronald J. Hedges, former United States Magistrate Judge for the District of New Jersey, for his contributions to this article.

[44] *Id.* at 14.

Disclaimer: The viewpoints expressed by the authors are their own and do not necessarily reflect the opinions, viewpoints and official policies of Practising Law Institute.