AI Ethics for Lawyers

A Practical Guide to Professional Responsibility in the Age of AI

Colin S. Levy
2026 Edition

About the Author

Colin S. Levy is a legal professional and commentator working at the intersection of law, technology, and organizational design. Over the course of his career, he has advised law firms, corporate legal departments, and legal technology companies on how to integrate emerging tools into the practice of law without sacrificing professional responsibility or analytical rigor. His perspective is shaped by direct involvement in legal operations, technology procurement, and the day-to-day realities of legal service delivery.

Colin is the author of the *Legal Technology Use Case Guide*, *AI for Lawyers: A Resource Guide*, the *AI Implementation Playbook for Legal Teams*, the *AI Agents Playbook for Legal Teams*, and *Legal Tech Resources: A Curated Guide*. He writes and speaks frequently on the ethics, governance, and practical mechanics of legal technology adoption.

This guide addresses a gap that has become increasingly apparent as AI adoption accelerates across the profession: the distance between the ethical rules that govern lawyers and the practical decisions lawyers face when using AI tools. The Model Rules of Professional Conduct were not drafted with generative AI in mind, but they apply to it. Bar authorities have begun issuing guidance, and the contours of a professional responsibility framework for AI use are taking shape. This guide translates that framework into specific, actionable protocols that legal teams can implement immediately.

Table of Contents

Part I

The Ethical Foundation

Professional obligation, applicable rules, and risk mapping

1. Why AI Ethics Is a Professional Obligation, Not a Philosophy Exercise

Conversations about AI ethics in the legal profession often begin in the wrong place. They start with abstract principles, hypothetical scenarios, or broad societal concerns about artificial intelligence. These discussions have value, but they are not where practicing lawyers need to begin. For lawyers, AI ethics is not a branch of philosophy. It is a set of concrete professional obligations, already codified in rules of professional conduct, already being interpreted by bar authorities, and already producing disciplinary consequences for those who ignore them.

The lawyers who have faced sanctions for AI-related conduct did not fail because they lacked a theory of machine ethics. They failed because they submitted AI-generated court filings without verifying the citations,[1] because they did not understand what their tools were doing, or because they treated AI output as a substitute for professional judgment rather than a starting point for it. The ethical failures that have emerged so far are failures of ordinary professional responsibility: competence, diligence, candor, and supervision.

The Practical Stakes

The consequences of ethical failures involving AI are not hypothetical. Courts have imposed monetary sanctions on attorneys who cited fabricated cases generated by AI tools.[1] Bar authorities have issued guidance making clear that existing disciplinary frameworks apply to AI use.[2] Clients have filed malpractice claims alleging that attorneys' reliance on unverified AI output constituted a breach of the standard of care. Insurance carriers are beginning to inquire about AI usage practices as part of underwriting.

These are not future risks. They are present realities. The question for every legal professional is not whether ethical obligations apply to AI use, but whether their current practices satisfy those obligations. This guide provides the framework and the specific protocols needed to answer that question with confidence.

What This Guide Does and Does Not Cover

This guide addresses the professional responsibility dimensions of AI use by practicing lawyers. It covers competence, confidentiality, supervision, candor, billing, and bias as they arise in the day-to-day use of AI tools for legal work. It translates bar authority guidance and Model Rule requirements into actionable protocols, checklists, and decision frameworks.

It does not address the broader societal ethics of artificial intelligence: questions about AI sentience, existential risk, or the philosophical foundations of machine decision-making. These are important questions, but they are not the questions that will determine whether a lawyer faces a disciplinary complaint next quarter. This guide focuses relentlessly on the practical.

> ***The Core Principle:*** *Every ethical obligation that applies to legal work performed by a human applies with equal force to legal work performed with the assistance of AI. The technology changes the tools available; it does not change the standard of care, the duty of confidentiality, or the obligation of candor. A lawyer who would not submit a brief without reading it cannot submit an AI-drafted brief without reading it. A lawyer who would not share client files with an unvetted third party cannot input client data into an unvetted AI system.*

2. The Rules That Already Apply: Model Rules and AI

The Model Rules of Professional Conduct were not drafted with AI in mind, but their principles are sufficiently broad to encompass it. The ABA's Formal Opinion 512, issued in July 2024, confirmed this interpretation and provided detailed guidance on how existing rules apply to generative AI tools.[2] Several state bars have issued their own guidance, each contributing distinct emphases. What follows is a practical translation of each relevant rule into specific obligations for AI-using lawyers.

Competence (Model Rule 1.1)

The competence obligation requires lawyers to provide "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation." Applied to AI, this means three things.

First, a lawyer must understand what the AI tool does at a functional level: what inputs it accepts, what outputs it produces, how it generates those outputs, and where it is known to fail. This does not require expertise in machine learning. It requires the same level of understanding that a lawyer would need of any tool relied upon for client work.

Second, a lawyer must understand the specific limitations relevant to legal work: the tendency of large language models to generate plausible but fabricated citations (commonly called hallucinations), the potential for outdated training data to produce legally incorrect analysis, and the inability of most AI tools to distinguish between binding and persuasive authority or to account for recent developments in the law.

Third, a lawyer must verify AI-generated output before relying on it. Every factual assertion must be independently confirmed. Every legal citation must be checked against primary sources. Every analytical conclusion must be evaluated through the lens of the lawyer's own professional judgment. ABA Formal Opinion 512 is explicit: uncritical reliance on AI-generated content, including citations, factual claims, and analytical conclusions, constitutes a competence failure.[2]

Confidentiality (Model Rule 1.6)

The confidentiality obligation prohibits a lawyer from revealing information relating to the representation of a client unless the client gives informed consent. Applied to AI, this creates a series of concrete requirements.

Before entering any client information into an AI tool, the lawyer must determine: whether the tool stores user inputs; whether those inputs are used to train or refine the model; whether the vendor's employees or subcontractors can access the data; whether the data traverses jurisdictions with different privacy regimes; and whether the tool's terms of service contain provisions that could compromise confidentiality. If the answers to these questions are unknown, the tool should not be used for client data until they are resolved.

ABA Formal Opinion 512 specifically extends the confidentiality analysis to the AI context: lawyers must evaluate the AI tool's data handling practices with the same rigor applied to any third-party service provider who might access client information.[2] Where a tool does not offer adequate confidentiality protections, its use on client matters is ethically precluded absent informed client consent.

Supervision (Model Rules 5.1 and 5.3)

Partners and supervising attorneys have an obligation to ensure that subordinate lawyers and non-lawyer assistants comply with professional conduct rules. ABA Formal Opinion 512 treats AI tools analogously to non-lawyer assistants: attorneys with supervisory authority bear responsibility for ensuring that AI is used appropriately within their organizations.[2]

In practical terms, this means that a supervising attorney who permits associates to use AI tools without providing guidance on verification, confidentiality, and quality control has failed to meet supervisory obligations. The obligation extends to establishing training programs, usage policies, and review protocols. It is not enough to issue a memorandum permitting AI use; the supervising attorney must create the conditions for responsible use and monitor compliance.

Candor Toward the Tribunal (Model Rules 3.1 and 3.3)

A lawyer must not make false statements of fact or law to a tribunal and must not offer evidence the lawyer knows to be false. AI-generated content that has not been independently verified poses a direct risk to this obligation. A lawyer who submits a brief containing AI-generated citations without checking them against primary sources risks presenting fabricated authority to the court.

Several courts have now imposed standing orders requiring attorneys to disclose when AI tools were used in the preparation of filings.[3] Even where no such order exists, the candor obligation requires that a lawyer not present AI-generated research as the product of independent legal analysis when it has not been independently verified. The standard is not whether the lawyer intended to mislead, but whether the filing is accurate.

Communication (Model Rule 1.4)

Lawyers must keep clients reasonably informed about the status of their matter and explain matters to the extent reasonably necessary for clients to make informed decisions. Applied to AI, this means that lawyers should communicate with clients about how AI supports their matters when AI use is material to the representation. This includes situations where AI tools are used for substantive legal analysis, where the use of AI materially affects the cost or timeline of the representation, or where the client has expressed preferences about AI use.

ABA Formal Opinion 512 does not require disclosure of every instance of AI use (just as lawyers do not disclose every use of Westlaw or spell-check), but it does require communication when AI use is material to the representation and when the client would reasonably want to know.[2]

> ***State Variations:*** *While the ABA Model Rules provide the national framework, individual states have issued guidance with varying emphases. Florida requires thorough review and independent fact-checking of all AI-generated documents before court submission (Bar Opinion 24-1, January 2024).[4] California organizes obligations around six core duties and requires a risk evaluation of each AI technology before deployment (Practical Guidance, November 2023).[5] Texas establishes four primary obligations including independent verification of all AI-generated information (Bar Opinion 705, February 2025).[6] New York addresses the specific scenario of using AI to record, transcribe, and summarize client conversations, imposing a consent standard stricter than the state's one-party consent law (Bar Opinion 2025-6).[7] Legal teams should monitor guidance from every jurisdiction in which they practice.*

3. Mapping Ethical Risk Across Legal AI Use Cases

Different uses of AI in legal practice carry different ethical risk profiles. A lawyer using AI to format a document faces negligible ethical risk. A lawyer using AI to draft a motion for summary judgment faces substantial risk across multiple dimensions: competence, candor, and potentially confidentiality. This chapter maps the ethical risk landscape across common legal AI use cases, providing a framework for calibrating the level of ethical diligence to the level of risk.

Lower-Risk Use Cases

Document Formatting and Administrative Tasks

Using AI to format documents, generate tables of contents, correct typographical errors, or organize files carries minimal ethical risk. The output is mechanical, errors are easily detected, and client data exposure is typically limited. Standard confidentiality precautions apply (no client data in unapproved tools), but the verification burden is light.

Internal Brainstorming and Outlining

Using AI to generate outlines, brainstorm arguments, or explore analytical frameworks for internal use carries moderate but manageable risk. The key safeguard is treating the output as a starting point, not a work product. If the AI suggests a line of argument, the lawyer must independently evaluate its merit and research its legal basis. The confidentiality risk depends on whether client-specific information is included in the prompts.

Moderate-Risk Use Cases

First-Draft Document Generation

Using AI to generate first drafts of contracts, memoranda, or correspondence creates moderate risk across several dimensions. The competence risk lies in the possibility that the draft contains substantive errors that a superficial review might miss: incorrect legal standards, inapplicable precedent, or provisions that conflict with the client's objectives. The confidentiality risk depends on how much client information is provided to the AI tool. The supervision risk is that junior attorneys may treat the AI draft as substantially complete when it requires significant revision.

Contract Review Against Standard Playbooks

AI-assisted contract review, where the tool compares clauses against a predefined playbook, creates moderate risk that centers on the completeness and accuracy of the tool's analysis. The ethical obligation is to verify that the tool has identified all material deviations, not just the most obvious ones. This requires that the reviewing attorney have sufficient expertise to recognize what the tool may have missed, which in turn

requires the competence to understand the tool's limitations.

Higher-Risk Use Cases

Legal Research and Citation

AI-assisted legal research carries high ethical risk because of the documented tendency of large language models to generate fabricated citations. A single fabricated citation in a court filing can result in sanctions, reputational damage, and disciplinary proceedings. The ethical obligation is absolute: every citation must be independently verified against primary sources. No exception exists for citations that "look right" or come from a tool that is "usually accurate." The verification must confirm that the case exists, that it remains good law, and that it stands for the proposition for which it is cited.

Client-Facing Communications

Using AI to draft communications that will be sent to clients, opposing counsel, or courts without substantial human revision carries high risk. The communication represents the lawyer's professional judgment, and errors in AI-generated communications are attributed to the lawyer. Beyond accuracy, the lawyer must ensure that the tone, strategy, and substance of the communication reflect informed professional judgment about the specific circumstances of the matter.

Predictive Analytics and Case Strategy

Using AI tools that predict case outcomes, judicial behavior, or settlement values to inform strategy decisions carries high risk because of the potential for bias in the underlying data and the opacity of the analytical models. The competence obligation requires the lawyer to understand the basis for the tool's predictions, the limitations of its training data, and the circumstances under which its predictions are unreliable. Relying on a prediction without this understanding is not a strategic decision; it is a delegation of judgment to a system the lawyer does not comprehend.

> ***Risk Calibration Protocol:*** *For each AI use case, assess risk along four dimensions: (1) the probability of substantive error in the AI output, (2) the detectability of that error during review, (3) the potential consequences if the error reaches the client, court, or counterparty, and (4) the confidentiality exposure inherent in the task. Use cases that are high on multiple dimensions require the most rigorous verification, supervision, and documentation protocols.*

Part II

Confidentiality and Data Protection

Client data flows, privilege considerations, and vendor obligations

4. Client Data in AI Systems: What Flows Where

Confidentiality is the ethical obligation most directly affected by AI adoption. Every time a lawyer enters information into an AI tool, data moves from the lawyer's control into a system operated by a third party. Understanding precisely what data flows where, and what happens to it after it arrives, is not a technical curiosity. It is a professional responsibility requirement.

Data Flow Analysis for Common AI Tools

Before using any AI tool for client work, conduct a data flow analysis that answers the following questions with specificity:

- What data does the tool receive? This includes not only the text of the prompt but metadata such as the user's identity, the time of the query, and any files attached or referenced.
- Where is the data transmitted? Identify the cloud provider, the geographic location of the servers, and whether data crosses jurisdictional boundaries that trigger additional privacy obligations.
- How long is the data retained? Distinguish between session-based retention (data deleted after the session ends), fixed-term retention (data stored for a defined period), and indefinite retention.
- Who can access the data? Determine whether the vendor's employees, contractors, or subprocessors can view user inputs, and under what circumstances.
- Is the data used for model training? This is the single most important question for legal confidentiality. If user inputs are used to train or improve the model, client information could theoretically surface in outputs provided to other users.

Practical Data Classification Protocol

Not all data carries the same confidentiality burden. Establish a classification system that governs what information may be entered into which tools:

Tier 1: Public Information

Publicly available information (published case law, statutes, public filings) may be entered into any approved AI tool without restriction. No confidentiality risk attaches to information that is already in the public domain.

Tier 2: Internal Work Product

Internal analysis, draft arguments, research notes, and strategic assessments that do not contain client-identifying information may be entered into AI tools that meet baseline security requirements: no training on user data, encrypted transmission, and contractual data handling commitments.

Tier 3: Confidential Client Information

Information that identifies clients or relates to specific matters requires the highest tier of protection. This data may be entered only into AI tools that have been specifically vetted for legal confidentiality: enterprise-grade platforms with no-training guarantees, SOC 2 Type II certification, data residency commitments, and contractual provisions that satisfy the lawyer's confidentiality obligations.

Tier 4: Privileged Communications

Attorney-client privileged communications and attorney work product should not be entered into third-party AI tools unless the platform has been evaluated for privilege implications and the organization has concluded that the use does not constitute a waiver. This assessment requires legal analysis, not merely technical assurance. Chapter 5 addresses privilege in detail.

> ***Practical Step:*** *Create a one-page reference card for each approved AI tool in your organization, specifying which data tiers it is cleared for, what confidentiality safeguards are in place, and what restrictions apply. Distribute these cards to every user and require acknowledgment. A policy that exists only in a governance document is a policy that will be violated by the first attorney who needs a quick answer on a Friday afternoon.*

5. Privilege, Work Product, and AI-Assisted Communications

The intersection of AI tools and legal privilege is one of the most consequential and least resolved areas of AI ethics for lawyers. When a lawyer enters privileged communications into a third-party AI tool, does that constitute a disclosure that waives the privilege? The answer depends on the specific circumstances, but the analysis must be conducted before the data is entered, not after.

The Waiver Analysis

Attorney-client privilege can be waived by voluntary disclosure to a third party. The critical question is whether the AI vendor constitutes a third party for privilege purposes and, if so, whether any exception applies. Two analytical frameworks are relevant.

The Kovel Doctrine and Functional Equivalents

Under the Kovel doctrine, communications shared with a third party who is functioning as an agent of the attorney for purposes of providing legal services may retain their privileged character. Some commentators have argued that AI tools function as agents of the attorney in this sense. However, this argument has not been tested in reported decisions, and its application is uncertain. The conservative approach is to assume that privilege may be at risk when privileged communications are shared with a third-party AI vendor and to act accordingly.

Reasonable Precautions and Inadvertent Disclosure

Under Federal Rule of Evidence 502(b), inadvertent disclosure of privileged material does not constitute a waiver if the holder took reasonable steps to prevent disclosure and promptly took reasonable steps to rectify the error. For AI use, this creates a framework: if the organization has established clear policies about what data may be entered into which tools, has trained its personnel, and has technical safeguards in place, an inadvertent entry of privileged data into an AI tool may not waive the privilege, provided the error is identified and addressed promptly. But relying on the inadvertent disclosure exception is a fallback, not a strategy.

Practical Protocols for Privilege Protection

- Default rule: privileged communications should not be entered into third-party AI tools unless the tool has been specifically cleared for privileged data through a documented legal analysis.
- If an AI tool is cleared for privileged data, document the basis for the conclusion: the contractual protections in place, the technical safeguards, and the legal analysis supporting the conclusion that use does not constitute waiver.

- Train all personnel to recognize privileged material and to apply the correct data classification before entering information into any AI tool.
- Establish a remediation protocol for situations where privileged data is inadvertently entered into an unapproved tool: immediate notification to the ethics officer or managing partner, a request to the vendor for data deletion, and documentation of the steps taken to mitigate any waiver argument.

Work Product Considerations

The work product doctrine protects materials prepared in anticipation of litigation. AI-generated drafts, research summaries, and analytical memoranda prepared in anticipation of litigation are likely protected as work product. However, the underlying prompts (which may reveal litigation strategy) and the selection of which AI outputs to rely on (which reflects attorney mental processes) may receive heightened protection as opinion work product. The practical implication is that the organization's data governance framework should protect not only the AI outputs but the prompts and the decision-making process surrounding them.

> ***Open Question:*** *As of early 2026, no appellate court has issued a definitive ruling on whether sharing privileged communications with a third-party AI vendor constitutes waiver. The prudent approach is to treat this as an unresolved risk and to err on the side of caution. Organizations that establish robust classification and access controls now will be well positioned regardless of how courts ultimately resolve the question.*

6. Vendor Diligence as an Ethical Obligation

Selecting an AI vendor is not merely a procurement decision; it is an ethical act. A lawyer who entrusts client data to an AI vendor without conducting adequate diligence has potentially breached the duty of confidentiality, regardless of whether a data incident actually occurs. The standard is not whether harm resulted, but whether the lawyer took reasonable steps to protect client information before the data was shared.

Minimum Diligence Requirements

The following vendor diligence steps should be treated as mandatory for any AI tool that will process client data:

- Obtain and review the vendor's SOC 2 Type II report. Verify that the audit is current (within the past twelve months) and that no material exceptions were noted. A Type I report, which evaluates controls at a point in time rather than over a period, is insufficient.
- Obtain written confirmation that user inputs and uploaded documents are not used to train, fine-tune, or improve the vendor's models. This confirmation should be contractual, not merely a statement in the vendor's privacy policy, which can be amended unilaterally.
- Review the vendor's data retention and deletion policies. Determine how long data is retained, under what circumstances it can be deleted, and whether deletion is certified in writing.
- Identify all subprocessors who may access client data and evaluate their security posture. A vendor that engages subprocessors without disclosure creates uncontrolled exposure.
- Verify that the vendor's data residency practices comply with applicable jurisdictional requirements, including any cross-border transfer restrictions that apply to the client's data.
- Review the vendor's incident response protocol: how breaches are detected, how quickly customers are notified, and what remediation is provided.

Contractual Protections

Beyond diligence, the engagement with the AI vendor should be documented in a contract that includes specific protections for legal confidentiality:

- A no-training clause prohibiting the use of customer data to train, improve, or develop the vendor's models, with a contractual remedy for breach.
- Data handling obligations specifying encryption standards (AES-256 at rest, TLS 1.3 in transit), access controls, and audit rights.
- Notification obligations requiring the vendor to inform the organization within a defined timeframe (24 to 72 hours is market standard) of any security incident that may affect client data.

• Termination and data deletion provisions requiring certified destruction of all customer data within a defined period following contract termination.

• Indemnification provisions allocating liability for data breaches caused by the vendor's failure to comply with its contractual obligations.

> ***Diligence as Defense:*** *Documented vendor diligence serves a dual purpose. It protects client data by ensuring that the vendor meets appropriate security standards. And it protects the lawyer by demonstrating that reasonable steps were taken to safeguard confidentiality. If a data incident occurs despite adequate diligence, the documented process provides evidence that the lawyer satisfied the standard of care. If diligence was not conducted, the lawyer has no defense regardless of the outcome.*

Part III

Competence, Supervision, and Candor

Practical standards for AI-assisted legal work

7. The Competence Standard for AI-Using Lawyers

The duty of competence has always required lawyers to understand their tools. The Comment to Model Rule 1.1 was amended in 2012 to include an obligation to stay current with "the benefits and risks associated with relevant technology."[8] This amendment, originally adopted in the context of e-discovery and electronic communications, now applies with particular force to AI tools, which are more capable, more opaque, and more prone to subtle failure than any technology lawyers have previously used.

What Competence Requires

Competence in the AI context does not require a lawyer to become a computer scientist. It requires a working understanding of five dimensions:

1. How the Tool Generates Output

The lawyer should understand, at a functional level, the type of AI system being used (large language model, retrieval-augmented generation, classification system), how it produces output (probabilistic text generation, database retrieval, pattern matching), and what that process implies about the reliability of the output. A lawyer who uses a large language model for legal research without understanding that it generates text based on statistical patterns rather than legal reasoning lacks the competence to evaluate the output.

2. Known Failure Modes

Every AI tool has characteristic failure modes. Large language models hallucinate: they generate plausible but fabricated content, including case citations, statutory references, and factual claims. Contract review tools may miss non-standard provisions that fall outside their training distribution. Predictive analytics tools may reflect biases in their training data. The competent lawyer knows the failure modes specific to each tool in use and designs verification procedures to catch them.

3. Appropriate Use Cases

Competence requires the judgment to determine which tasks are appropriate for AI assistance and which are not. This judgment depends on the risk profile of the task (Chapter 3), the reliability of the tool for that task type, and the lawyer's ability to verify the output. A tool that performs well for first-draft generation may be unreliable for citation verification, and the competent lawyer does not generalize from one use case to another without independent evaluation.

4. Verification Methods

The competent lawyer has specific verification protocols for each AI use case: citation checking against primary sources for legal research, clause-by-clause review against the playbook for contract review, factual verification against source documents for summarization. These protocols should be documented, trained,

and enforced. Chapter 8 provides detailed verification frameworks.

5. When Not to Use AI

Perhaps the most important dimension of competence is the judgment to recognize when AI assistance is inappropriate: when the task requires nuanced professional judgment that the tool cannot provide, when the risk of undetected error is disproportionate to the efficiency gain, or when the confidentiality exposure exceeds what the available tools can safely manage. The competent lawyer uses AI as a tool, not a crutch, and maintains the independent professional judgment to override, supplement, or decline AI assistance as circumstances require.

> ***Competence Benchmark:*** *Ask yourself: if a disciplinary authority asked you to explain how this AI tool works, why you chose it for this task, what steps you took to verify its output, and how you determined that its use was appropriate given the circumstances, could you provide a clear, specific, and credible answer? If not, additional diligence is required before proceeding.*

8. Supervising AI-Assisted Work Product

The supervisory obligations under Model Rules 5.1 and 5.3 require attorneys with managerial or supervisory authority to establish policies and procedures that provide reasonable assurance of compliance with professional conduct rules. For AI-assisted work, this means creating organizational infrastructure that ensures every AI output is reviewed with appropriate rigor before it is relied upon, distributed, or filed.

A Verification Framework by Use Case

Different AI use cases require different verification protocols. The following framework specifies the minimum verification required for common AI applications in legal practice:

Legal Research

- Verify that every cited case exists by checking against the official reporter or a primary legal database.
- Confirm that each case remains good law (not reversed, overruled, or distinguished on the relevant point) using a citator service.
- Read each cited case to verify that it stands for the proposition attributed to it, not merely that it exists and is good law.
- Verify that the applicable standard of review, burden of proof, and jurisdictional rules are correct for the relevant forum.

Contract Drafting and Review

- Compare each AI-generated or AI-reviewed clause against the applicable playbook, template, or client instructions.
- Verify that defined terms are used consistently and that cross-references are accurate.
- Confirm that the contract reflects the commercial terms agreed upon by the parties, not generic provisions imported from the AI's training data.
- Review for jurisdiction-specific requirements (governing law, dispute resolution, regulatory compliance provisions) that the AI may have applied incorrectly.

Client Communications

- Verify all factual statements against the underlying matter file.
- Ensure that legal advice reflects the lawyer's independent professional judgment, not merely the AI's output.
- Review tone and strategy to confirm that the communication serves the client's specific interests in the specific circumstances.

Institutional Supervision Structures

Individual verification is necessary but not sufficient. Organizations must also establish structural safeguards:

- Designate a responsible attorney (or committee) with authority over AI usage policies, approved tool lists, and compliance monitoring.
- Conduct periodic quality audits of AI-assisted work product, sampling outputs across practice groups and seniority levels to identify patterns of error or insufficient review.
- Maintain incident logs documenting AI errors that reached or nearly reached clients, courts, or counterparties. Analyze these incidents for root causes and systemic remedies.
- Require that AI usage be documented in matter files so that reviewing attorneys and successors can identify which work product was AI-assisted and evaluate it accordingly.

> ***The Supervision Standard:*** *Review AI-generated work product with the same skepticism you would apply to a draft from a first-year associate who is intelligent but unfamiliar with the specific area of law: assume the structure will be reasonable but the details may be wrong, the citations may need checking, and the analysis may lack the nuance that comes from experience with the particular client, jurisdiction, or counterparty.*

9. Disclosure, Candor, and Transparency Obligations

The question of when and how lawyers must disclose their use of AI tools is evolving rapidly. No universal disclosure requirement exists, but obligations arise from multiple sources: court standing orders, client engagement terms, bar authority guidance, and the general duties of candor and communication. This chapter maps the current disclosure landscape and provides practical protocols for each context.

Disclosure to Courts

A growing number of federal and state courts have issued standing orders or local rules requiring attorneys to disclose the use of AI tools in the preparation of court filings.[3] These requirements vary in scope: some require disclosure of any AI use, others are limited to generative AI, and others require certification that all AI-generated content has been independently verified. Regardless of whether a specific court has issued such an order, the duty of candor under Model Rule 3.3 requires that attorneys not present AI-generated work as independently researched legal analysis when it has not been independently verified.

The practical protocol is straightforward: before filing in any court, check whether a standing order or local rule addresses AI disclosure. If one exists, comply with its specific requirements. If none exists, ensure that the filing is accurate, that all citations have been verified, and that the analysis reflects the attorney's independent professional judgment. Document the verification steps taken.

Disclosure to Clients

ABA Formal Opinion 512 does not require disclosure of every instance of AI use, but it does require communication when AI use is material to the representation.[2] Materiality should be assessed on a case-by-case basis, but the following situations presumptively require disclosure: when AI is used for substantive legal analysis that forms the basis of advice to the client; when AI use materially affects the cost of the representation (requiring an adjustment to the fee arrangement); when the client has expressed a preference regarding AI use; and when the engagement letter or organizational policy requires it.

Consider addressing AI use proactively in engagement letters. A simple paragraph explaining that the firm may use AI tools to support certain aspects of the representation, subject to attorney review and quality control, provides transparency without creating unnecessary alarm. Some firms have adopted this approach as standard practice; others address it on a matter-by-matter basis as warranted.

What Informed Consent Actually Looks Like

ABA Formal Opinion 512 requires more than boilerplate when client data will be entered into self-learning AI tools. Informed consent under Model Rule 1.6(a) requires that the client understand the material risks of the disclosure and the alternatives to it. Applied to AI, this means communicating: what types of AI tools

will be used and for what purposes; how client data is handled by those tools (storage, training, access); what safeguards are in place to protect confidentiality; how AI-assisted work is reviewed and verified by attorneys; and how AI use may affect the cost or timeline of the engagement. A generalized statement that the firm "may use technology" is insufficient. The disclosure must be specific enough for the client to make a meaningful decision.

Engagement Letter Language: Substantive AI Use

For matters where AI will be used for substantive legal work (research, drafting, analysis), consider language along these lines: "In the course of this engagement, our attorneys may use AI-assisted tools to support legal research, document review, contract analysis, and initial drafting. All AI-generated content is reviewed and verified by a licensed attorney before being relied upon or delivered as work product. The AI platforms we use are subject to our firm's data security and confidentiality protocols: your information is processed under enterprise agreements that prohibit the use of client data for model training, require encryption in transit and at rest, and restrict access to authorized personnel. AI tools may enable us to work more efficiently, and any resulting time savings are reflected in our billing practices. If you have questions about our use of AI tools on your matter, or if you prefer that we limit or exclude their use, please let us know. Your signature on this engagement letter constitutes your acknowledgment of and consent to our use of AI as described above."

Engagement Letter Language: Non-Substantive AI Use

For matters where AI use is limited to administrative or formatting tasks, a lighter disclosure is appropriate: "Our firm may use AI-assisted tools for administrative tasks such as document formatting, scheduling, and proofreading. These tools do not perform substantive legal analysis, and their use does not affect the attorney oversight applied to your matter."

Key Principles for Effective Consent Language

- Be specific about what the tools do, not just that they exist. Clients cannot consent meaningfully to something described only in vague terms.
- Address data handling directly. The client's primary concern is typically whether their confidential information is being shared with or stored by a third party.
- Explain the human review layer. Clients need to understand that AI is assisting, not replacing, their attorney's professional judgment.
- Connect AI use to billing. If AI reduces the time required for a task, the client should understand how that efficiency is reflected in what they pay.
- Preserve the client's right to opt out. Informed consent requires the ability to say no. If a client objects to AI use on their matter, accommodate the preference and document it.

> ***Beyond the Engagement Letter:*** *The engagement letter provides baseline consent, but informed consent is an ongoing obligation. If the firm adopts a new AI tool, changes vendors, or begins using AI for a purpose not contemplated at the outset of the engagement, the client should be informed and given the opportunity to consent to the new use. This is particularly important when the new tool processes more sensitive categories of client data than previously authorized.*

Disclosure to Opposing Counsel and Third Parties

No general obligation requires disclosure of AI use to opposing counsel or third parties. However, specific circumstances may create such an obligation: for example, if a discovery response was prepared using an AI tool that is known to have limitations relevant to the completeness of the response, candor may require disclosure of the methodology. Similarly, in transactional contexts where both parties are relying on the accuracy of a document, disclosure of AI involvement may be appropriate if the AI tool's limitations could affect the reliability of the work product.

> ***Practical Protocol:*** *Maintain a disclosure decision matrix that maps each disclosure context (courts, clients, opposing counsel, regulators) against each AI use case. For each combination, specify the default disclosure position and the criteria for departing from it. This matrix should be reviewed and updated as court orders, bar guidance, and organizational experience evolve.*

Part IV

Billing, Bias, and Accountability

Fee ethics, fairness obligations, and building a sustainable practice

10. Fee Ethics in an AI-Augmented Practice

AI tools can reduce the time required to complete legal tasks by significant margins. This efficiency creates an ethical obligation that multiple bar authorities have addressed with unusual directness: fees must reflect the actual effort expended. If an AI tool reduces a task from four hours to forty minutes, billing four hours is ethically impermissible.[2][4][6] The question is not whether fees must be adjusted, but how.

The Ethical Framework for AI-Affected Billing

Model Rule 1.5 requires that fees be reasonable. The factors relevant to reasonableness include the time and labor required, the novelty and difficulty of the questions involved, the skill requisite to perform the legal service properly, and the results obtained. AI tools affect at least the first of these factors directly and may affect the others indirectly.

Hourly Billing

For matters billed on an hourly basis, the ethical obligation is to bill for the time actually spent, including both AI-assisted execution time and attorney review time. If an attorney uses AI to generate a first draft in fifteen minutes and then spends forty-five minutes reviewing, revising, and verifying the output, the billable time is one hour, not the four hours the task would have required without AI assistance. The attorney may not inflate time entries to recapture efficiency gains.

Flat Fees and Alternative Fee Arrangements

For matters billed on a flat-fee basis, the ethical analysis is different. A flat fee represents the value of the service, not the time required to perform it. If AI enables the lawyer to deliver the same quality of work in less time, the flat fee remains reasonable provided it was fair at inception and the quality of service is maintained. However, if AI dramatically reduces the cost of delivery, market forces and the duty of communication may require the lawyer to adjust fees over time to reflect the new economics of the engagement.

Cost Pass-Through

ABA Formal Opinion 512 addresses whether AI tool costs may be passed to clients. A lawyer may treat AI costs as part of office overhead, or may charge clients for AI tool usage on a per-use basis, provided the arrangement is disclosed to the client in advance and informed consent is obtained.[2] The opinion also notes that a lawyer may not charge a client for the time spent learning to use a technology that will be used for multiple clients, unless the client specifically requested the use of that particular tool.

Practical Billing Protocols

- Record time entries with sufficient specificity to distinguish between AI-assisted and manual work. Time entries that read "research and drafting" without indicating AI involvement create audit risk.

• Establish internal guidelines for how AI-assisted tasks are recorded: the total time spent (including review), the nature of the AI assistance, and the attorney who reviewed the output.

• Review AI-affected time entries before invoicing to ensure they reflect the actual effort expended and do not inadvertently bill for time saved by AI assistance.

• Address AI billing practices in engagement letters, specifying whether AI costs are included in hourly rates, billed separately, or treated as overhead.

> ***The Billing Test:*** *Before submitting any invoice that includes AI-assisted work, ask: if the client knew exactly how this work was performed, including the role of AI and the time actually spent, would they consider the fee reasonable? If the answer is uncertain, the time entry needs revision.*

11. Bias, Fairness, and the Duty of Diligent Representation

AI systems can reflect and amplify biases present in their training data. For lawyers, this is not merely a theoretical concern about algorithmic fairness; it is a professional responsibility issue that intersects with the duties of competence and diligent representation. A lawyer who relies on a biased AI tool without understanding or accounting for that bias may be providing representation that falls below the standard of care.

Where Bias Manifests in Legal AI

Predictive Analytics and Case Assessment

AI tools that predict case outcomes, assess litigation risk, or recommend settlement values are trained on historical data that reflects the decisions of judges, juries, and opposing counsel. If those historical decisions were influenced by racial, gender, or socioeconomic bias, the AI tool will reproduce that bias in its predictions. A risk assessment tool that systematically undervalues claims brought by plaintiffs of a particular demographic is not providing neutral analysis; it is encoding discrimination.

Legal Research and Case Selection

AI research tools may prioritize certain types of cases or legal theories over others based on the frequency with which they appear in the training data. This can create blind spots: novel legal theories, emerging areas of law, or arguments that are sound but underrepresented in existing databases may be systematically underweighted. The competent lawyer recognizes that AI research is comprehensive only within the bounds of its training data and supplements it with independent analysis.

Contract Language and Drafting

AI drafting tools may generate provisions that reflect the norms of the jurisdictions and contract types most represented in their training data. This can produce drafts that are poorly suited to transactions involving parties in underrepresented jurisdictions, non-standard deal structures, or industry-specific requirements. The reviewing attorney must evaluate AI-generated drafts for appropriateness to the specific transaction, not merely for internal consistency.

Practical Steps to Mitigate Bias Risk

- Ask vendors specifically how they test for bias in their legal AI products: what evaluation metrics they use, what demographic groups they test across, and what results they have obtained. Vendors that cannot answer these questions have not adequately addressed the issue.
- Conduct your own testing when feasible: run the same analysis with and without demographic identifiers and compare results. Significant differences warrant investigation.

• Maintain awareness that AI tools are optimized for the majority of their training distribution. Edge cases, novel theories, and underrepresented contexts require heightened human oversight.

• Document your awareness of bias risks and the steps taken to mitigate them for each AI tool in use. This documentation serves both as an internal quality measure and as evidence of diligence if the tool's bias is later called into question.

> ***The Diligence Standard:*** *A lawyer need not eliminate all bias from AI tools (an impossible task). The obligation is to be aware of the potential for bias, to take reasonable steps to identify and mitigate it, and to exercise independent professional judgment rather than deferring uncritically to AI outputs that may reflect embedded biases. Awareness and active mitigation, not perfection, is the standard.*

12. Building an Ethical AI Practice: Protocols and Checklists

The principles and frameworks described in this guide must be translated into daily practice. Ethical AI use does not happen by default; it happens because an organization builds the structures, trains the people, and enforces the standards that make it possible. This chapter provides the operational building blocks.

The Pre-Use Checklist

Before using any AI tool for client work, the attorney should confirm:

- The tool is on the organization's approved list and has been cleared for the data classification of the information being entered.
- The task is appropriate for AI assistance given its risk profile (Chapter 3).
- The attorney understands the tool's capabilities, limitations, and known failure modes for this specific use case.
- A verification protocol exists for the type of output the tool will generate, and the attorney has allocated time to execute it.
- No privileged or Tier 4 data will be entered into a tool that has not been specifically cleared for privileged material.

The Post-Output Checklist

After receiving AI output, the attorney should confirm:

- Every factual assertion has been independently verified against primary or authoritative sources.
- Every legal citation has been checked for existence, current validity, and accuracy of the attributed proposition.
- The analysis reflects the attorney's independent professional judgment, not merely the AI's output rephrased.
- The output is appropriate for the specific client, matter, jurisdiction, and audience.
- The time entry accurately reflects the time actually spent, including both AI-assisted execution and attorney verification.

Organizational Infrastructure

AI Ethics Policy

Draft and adopt a written policy that addresses: approved tools and their data classifications, acceptable use cases and prohibited uses, verification requirements by use case, supervision and review standards,

disclosure obligations, billing practices, and incident reporting procedures. Review the policy at least annually and update it in response to new bar guidance, court orders, or organizational experience.

Training Program

Develop a training program that covers the ethical obligations described in this guide, the organization's specific policies and protocols, hands-on practice with approved tools, and exercises in identifying common AI failure modes. Require completion for all attorneys and staff who use AI tools. Refresh training when tools, policies, or regulatory guidance change.

Incident Response Plan

Establish a defined process for responding to AI-related ethical incidents: an AI-generated filing that contains errors, a confidentiality exposure through an AI tool, a client complaint about AI-assisted work, or a court inquiry about AI use. The plan should identify who is responsible for triage, what documentation is required, when escalation to firm leadership or ethics counsel is triggered, and how lessons learned are incorporated into policies and training.

Continuous Monitoring

Ethical AI practice is not a one-time implementation; it is an ongoing commitment. Monitor bar publications and ethics hotline advisories from every jurisdiction in which the organization practices. Track court orders regarding AI disclosure. Review vendor security posture and data handling practices periodically, not just at contract inception. Audit AI-assisted work product regularly. Update policies and training in response to what you learn.

Professional Liability Insurance and AI-Related Errors

A question that many firms have not yet confronted is whether their professional liability insurance actually covers claims arising from AI-related errors. The short answer, as of early 2026, is that coverage is uncertain and evolving, and firms that assume they are protected may be unpleasantly surprised when a claim is presented.[13]

The Coverage Gap

Most legal professional liability (LPL) policies were drafted before generative AI entered legal practice. They were designed to cover errors arising from the exercise of professional judgment by a licensed attorney. Several features of AI-assisted work create potential coverage gaps. First, if a lawyer relies on AI output without meaningful review, an insurer may argue that no "professional service" was provided, because the lawyer delegated the substantive work to a machine rather than exercising professional judgment. No professional service means no coverage under a standard LPL policy. Second, feeding confidential client information into an unsecured AI tool could be treated as an intentional disclosure rather than a negligent act. Most LPL policies exclude coverage for intentional breaches of confidentiality. Third, if an AI tool effectively makes legal judgments without attorney oversight, this could be characterized as the unauthorized practice of law, which is a standard policy exclusion.

Emerging AI-Specific Exclusions

Some insurers have begun attaching AI-specific exclusions to professional liability policies. These endorsements remove coverage for any claim involving the actual or alleged use of generative AI, sometimes defined broadly enough to encompass any system that produces text, imagery, or synthetic data in response to prompts. Firms should review their current and renewal policies carefully for such endorsements, which may appear as manuscript endorsements rather than standard policy language.

When Coverage Is More Likely to Apply

Insurers and underwriters have indicated that when firms use AI appropriately, with documented oversight, verification protocols, and attorney review, coverage is more likely to extend to claims that include an AI-related component. The key variable is whether the lawyer can demonstrate that professional judgment was exercised. An attorney who uses AI to generate a first draft, reviews it thoroughly, verifies all citations, and applies independent analysis is still providing a professional service. An attorney who submits AI output without review is not.

Practical Steps for Firms

- Review your current LPL policy for AI-related exclusions or limitations. Ask your broker or carrier directly whether claims arising from AI-assisted work are covered.
- Document your AI governance practices, including verification protocols, training programs, and approved tool lists. This documentation strengthens the argument that AI-assisted work constitutes a professional service subject to attorney oversight.
- Consider whether your cyber liability insurance provides supplemental coverage for AI-related data breaches or confidentiality exposures that may fall outside your LPL policy.
- As renewal periods approach, inquire about AI-specific coverage options. The insurance market is developing new products to address AI risk, and early engagement with your carrier positions you to secure appropriate coverage.
- Maintain incident logs documenting any AI errors caught during review. This record demonstrates that your verification protocols are functioning, which supports both coverage arguments and underwriting assessments.

> ***The Insurance Imperative:*** *The same practices that satisfy your ethical obligations, including attorney review, verification protocols, documentation, and oversight, are the practices that preserve your insurance coverage. An insurer evaluating whether a "professional service" was provided will look for evidence that a lawyer exercised professional judgment. Firms that can demonstrate robust AI governance are better positioned both ethically and from an insurance perspective. The two concerns are not separate; they are the same concern viewed from different angles.*

> ***The Institutional Commitment:*** *The organizations that will navigate AI ethics most successfully are those that treat it not as a compliance burden but as a dimension of professional excellence. An organization that can demonstrate robust AI governance, rigorous verification protocols, and a culture of responsible innovation will earn client trust, attract talent, and reduce risk. The investment in ethical infrastructure pays dividends that extend well beyond regulatory compliance.*

Appendix: Ethical AI Self-Assessment Checklist

The following checklist is designed for legal teams assessing the adequacy of their ethical framework for AI use. Each item represents a practice or capability that, if absent, creates ethical exposure. The checklist can be used for initial assessment, periodic review, or preparation for a specific AI deployment.

Policy and Governance

- A written AI ethics policy has been adopted, addressing approved tools, acceptable use, data classification, verification requirements, supervision standards, disclosure obligations, and billing practices.
- The policy incorporates current guidance from the ABA (Formal Opinion 512) and from every state bar in which the organization practices.
- A responsible individual or committee has been designated with authority over AI ethics policy, tool approval, and compliance monitoring.
- The policy is reviewed and updated at least annually or whenever significant new guidance is issued.

Confidentiality and Data Protection

- A data classification framework (Tiers 1 through 4) governs what information may be entered into which AI tools.
- Vendor diligence has been completed for every AI tool used for client work, including SOC 2 Type II review, no-training confirmation, and data handling assessment.
- Contractual protections (no-training clause, encryption requirements, breach notification, data deletion provisions) are in place with every AI vendor.
- A privilege analysis has been conducted for any AI tool into which privileged communications may be entered.
- Reference cards specifying data classification clearances are distributed to all AI tool users.

Competence and Verification

- All attorneys who use AI tools have received training on tool capabilities, limitations, known failure modes, and verification protocols.
- Verification protocols are documented for each AI use case (legal research, contract review, drafting, client communications).
- A risk mapping framework has been applied to identify which AI use cases require heightened verification and supervision.

Supervision and Quality Control

- Supervisory attorneys have established AI usage policies, training requirements, and review protocols for their teams.
- Periodic quality audits of AI-assisted work product are conducted across practice groups.
- Incident logs are maintained for AI errors that reached or nearly reached clients, courts, or counterparties.

Disclosure, Transparency, and Informed Consent

- A disclosure decision matrix maps each disclosure context (courts, clients, opposing counsel) against each AI use case.
- Engagement letters contain specific AI disclosure language that addresses tools used, data handling practices, attorney review protocols, billing implications, and the client's right to opt out.
- Informed consent is obtained before entering client information into self-learning AI tools, going beyond generalized boilerplate to address the specific risks and safeguards applicable to the engagement.
- Court standing orders regarding AI disclosure are monitored and compiled for every jurisdiction in which the organization files.
- A process exists to notify clients and obtain updated consent when the firm adopts new AI tools or changes vendors mid-engagement.

Billing and Fees

- Time-keeping practices distinguish between AI-assisted and manual work and reflect actual effort expended.
- Billing guidelines address how AI efficiency gains are reflected in client invoices.
- Engagement letters specify how AI tool costs are handled (overhead, pass-through, or separate charge).

Insurance and Risk Transfer

- The firm's professional liability policy has been reviewed for AI-related exclusions or limitations, and the carrier has confirmed the scope of coverage for AI-assisted work.
- Cyber liability insurance has been evaluated as supplemental coverage for AI-related data breaches or confidentiality exposures.
- AI governance documentation (verification protocols, training records, approved tool lists) is maintained in a form that supports coverage arguments in the event of a claim.

Endnotes

1. See, e.g., *Mata v. Avianca, Inc.*, No. 22-cv-1461 (S.D.N.Y. June 22, 2023) (imposing $5,000 sanctions on attorneys who submitted a brief containing six fabricated case citations generated by ChatGPT); *Park v. Kim*, No. 22-2057 (2d Cir. Jan. 30, 2024) (referring attorney for sanctions after submitting brief citing a non-existent case generated by AI).

2. American Bar Association, Formal Opinion 512, "Generative Artificial Intelligence Tools" (July 29, 2024). The opinion addresses competence (Model Rule 1.1), confidentiality (Model Rule 1.6), communication (Model Rule 1.4), candor toward the tribunal (Model Rules 3.1 and 3.3), supervisory responsibilities (Model Rules 5.1 and 5.3), and reasonable fees (Model Rule 1.5).

3. For an evolving compilation of court orders addressing AI use in litigation, see the tracking resources maintained by legal technology commentators and bar associations. As of early 2026, standing orders or local rules addressing AI have been issued by judges in numerous federal district courts and several state courts.

4. The Florida Bar, Ethics Opinion 24-1 (January 19, 2024), requiring thorough review and independent fact-checking of all AI-generated documents before court submission.

5. State Bar of California, Standing Committee on Professional Responsibility and Conduct, "Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law" (November 16, 2023).

6. Professional Ethics Committee for the State Bar of Texas, Opinion 705 (February 2025), developed by the Taskforce for Responsible AI in the Law (TRAIL).

7. New York City Bar Association, Formal Opinion 2025-6, "Ethical Issues Affecting Use of AI to Record, Transcribe, and Summarize Conversations with Clients" (2025). See also New York State Bar Association, Report and Recommendations of the Task Force on Artificial Intelligence (approved by the House of Delegates, April 6, 2024), 85 pages.

8. The amendment to Comment [8] of Model Rule 1.1 was adopted by the ABA House of Delegates in 2012 as part of the recommendations of the ABA Commission on Ethics 20/20. The comment now reads, in relevant part: "To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with relevant technology*." As of early 2026, over 40 states have adopted this or substantially similar language.

9. Federal Rule of Evidence 502(b) provides that disclosure of a communication or information covered by the attorney-client privilege or work-product protection does not operate as a waiver if the disclosure is inadvertent and the holder took reasonable steps to prevent disclosure and promptly took reasonable steps to rectify the error.

10. For a discussion of the Kovel doctrine in the AI context, see United States v. Kovel, 296 F.2d 918 (2d Cir. 1961), establishing that communications with third parties assisting the attorney in providing legal services may retain their privileged character under certain circumstances.

11. Thomson Reuters, *The ROI of Legal Tech & AI* (2025); ABA Formal Opinion 512 (July 29, 2024), specifically addressing the ethics of AI cost allocation and fee reasonableness.

12. Research on automation bias indicates that the tendency to defer to automated systems increases with system sophistication and perceived authority. See Parasuraman & Manzey, "Complacency and Bias in Human Use of Automation: An Attentional Integration," *Human Factors*, vol. 52, no. 3, pp. 381-410 (2010).

13. See Mark Bassingthwaighte, "Insurance Coverage Issues for Lawyers in the Era of Generative AI," ALPS Insurance (August 2025), discussing coverage gaps in standard legal professional liability policies for AI-related claims. See

also ABA Journal, "Does Your Professional Liability Insurance Cover AI Mistakes? Don't Be So Sure" (2025), examining the emerging trend of AI-specific exclusions in professional liability endorsements.