Addressing AI Concerns

Colin S. Levy

2026

This document is for informational purposes only and does not constitute legal advice.

About the Author

Colin S. Levy is a legal technology advocate, writer, and advisor who works at the intersection of law, technology, and business. With experience spanning in-house legal roles, legal technology companies, and legal operations, he brings a practical perspective to how legal teams can adopt and govern emerging technologies responsibly.

Colin writes and speaks extensively on legal innovation, artificial intelligence in legal practice, and the evolving role of legal professionals in a technology-driven landscape. His work focuses on helping legal teams move beyond the hype cycle to make sound, informed decisions about the tools they use and the workflows they build.

He is the author of The Legal Tech Ecosystem and editor of the Handbook of Legal Tech, and a regular contributor to publications covering legal technology and operations. He advises organizations on responsible AI adoption, legal operations strategy, and the practical governance frameworks that make innovation sustainable.

This guide is part of a series on AI and law that includes Managing AI Hallucinations, AI for Lawyers, AI for Legal Teams, AI Agents Data Handling and Cybersecurity Guide, AI in the Courtroom, Contracting with AI Vendors, Human Judgment and AI in Legal Practice, and the AI Implementation Playbook for Legal Teams.

Table of Contents

Part One

Why This Guide Exists

The Five Sentences You Have Heard

Every lawyer who has watched the past three years of generative AI adoption has said, or heard, some version of these sentences:

- “I cannot put client data into something that trains on my inputs.”
- “I do not trust an output I cannot trace to a source.”
- “If I get sanctioned, my career is over.”
- “My malpractice carrier has not told me whether this is covered.”
- “Which rules even apply when the model is in California, the vendor in Ireland, and my client in Brazil?”

Each of these sentences is reasonable. Each rests on a real risk that has been documented in real cases, real bar opinions, real regulatory enforcement, and real insurance disputes. None of them is an argument against using AI in legal practice. They are arguments for using it inside a defensible system.

What This Guide Is, and Is Not

This guide takes seriously the concerns lawyers actually voice about using AI tools, validates each, explains the mechanism behind it, and gives the reader a framework to address it. It draws on the American Bar Association’s Formal Opinion 512, the published guidance of state bars in the United States, the Solicitors Regulation Authority in England and Wales, the Law Society of Ontario, the Singapore courts, the Supreme Court of New South Wales, the European AI Act, and the growing body of case law in which courts have addressed lawyer use of AI directly.[1]

It is not a survey of AI capabilities, a vendor comparison, or a prediction about the future of legal work. It is a guide to using AI without losing your license, your client, your insurance coverage, or your reputation. The standard it returns to throughout is the standard most professional regimes already apply to every other tool in the lawyer’s hand: what would a reasonable, competent practitioner have done.

How the Guide Is Organized

Eight Parts follow, each anchored to a concern lawyers raise out loud. Each ends with a Concern Resolved callout that states what you can do, or stop worrying about, having read it. A jurisdictional Part addresses the variation across major bars and regulatory regimes. The final Part presents a concerns-to-controls map and a maturity model so the reader can place their own practice on a spectrum.

Key Principle:

Every concern lawyers raise about AI is reasonable. Most have systemic answers. The lawyers who get hurt are not the ones who proceed cautiously. They are the ones who proceed without a system.

Part Two

“Will I lose my license?”

The Reasonable Lawyer Standard

Across nearly every common-law jurisdiction, and in much civil-law professional regulation, professional discipline turns on a single test: what would a reasonable, competent practitioner have done in the same circumstances. The standard predates AI by a century. It already covers AI. The risk of losing your license does not come from using AI. It comes from using AI in a way no reasonable lawyer would defend if asked to explain it under oath.[2]

Federal Rule of Civil Procedure 11 requires every signing attorney to certify that the legal contentions in a filing are warranted by existing law and that factual contentions have evidentiary support. Typing a citation into Westlaw, Lexis, or even Google Scholar takes seconds. Failing to do so after generating that citation with a tool known to fabricate is difficult to defend as reasonable inquiry under Rule 11.[3]

The Sanctions Track Record

The case law is now substantial enough to be tracked as its own database. Researcher Damien Charlotin maintains a public repository of legal decisions in which courts have addressed lawyer reliance on AI-generated content. As of early 2026 the database contains more than 1,300 documented cases drawn from courts in the United States, Canada, the United Kingdom, Australia, and elsewhere. The pace accelerated through 2025 to multiple new cases per day.[4] The pattern is consistent. The lawyer submits AI-generated content to a tribunal, fails to verify it, and is sanctioned, referred for discipline, or both.

Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. 2023), remains the leading case. Plaintiff’s counsel filed an affirmation citing six fabricated decisions generated by ChatGPT, then provided fabricated opinion texts when challenged. Judge Castel found subjective bad faith and imposed $5,000 in sanctions.[5] In Park v. Kim, 91 F.4th 610 (2d Cir. 2024), the Second Circuit referred counsel to its Grievance Panel for citing a nonexistent case generated by ChatGPT and making no inquiry into its validity.[6] In Noland v. Land of the Free, L.P. (Cal. Ct. App., 2d Dist., Sept. 2025), the California Court of Appeal published its first opinion on AI-fabricated authority, imposing a $10,000 monetary sanction and reporting counsel to the State Bar after finding that briefs were “replete” with fabrications generated using ChatGPT, Claude, Gemini, and Grok.[7] In Zhang v. Chen, 2024 BCSC 285, the Supreme Court of British Columbia ordered counsel to personally compensate opposing counsel for time spent unwinding fabricated authority, and the Law Society of British Columbia opened an investigation.[8]

Pre-Filing AI Verification Checklist

A working document for any filing in which AI was used at any stage. Initial each step. Retain in the matter file.

- ■ Every case citation, statute citation, and regulatory citation in the filing has been independently retrieved from Westlaw, Lexis, the official reporter, or the issuing agency.
- ■ Every quotation attributed to a source has been compared to the source text and matches.
- ■ Every factual assertion drawn from AI output has been confirmed against the underlying record or a primary source.
- ■ The signing attorney has personally read the filing in full and is prepared to defend each contention under Rule 11.
- ■ Any standing order or local rule on AI disclosure for this court has been reviewed and complied with.

Signed: _____ Date: _____

ABA Formal Opinion 512 in One Page

On July 29, 2024, the ABA Standing Committee on Ethics and Professional Responsibility issued Formal Opinion 512, the first comprehensive opinion on lawyer use of generative AI. It addresses six duties: competence (Rule 1.1), confidentiality (Rule 1.6), communication (Rule 1.4), candor (Rules 3.1 and 3.3), supervision (Rules 5.1 and 5.3), and reasonable fees (Rule 1.5).[9] The opinion does not require lawyers to become AI experts. It requires a reasonable understanding of the capabilities and limitations of any generative AI tool the lawyer uses. The opinion treats generative AI tools as nonlawyer assistants for supervision purposes, requiring policies, training, and verification protocols consistent with Rules 5.1 and 5.3.

Cross-Border Convergence

The convergence across jurisdictions is striking. The Solicitors Regulation Authority's Risk Outlook on AI in the legal market emphasizes accuracy, accountability, and the duty to understand the tools in use.[10] The Law Society of Ontario's April 2024 white paper identifies the same core obligations: competence, confidentiality, supervision, candor, and fees.[11] The Singapore Registrar's Circular No. 1 of 2024 holds that the user remains fully responsible for AI-generated output and may face costs sanctions, evidentiary discounts, or disciplinary action for non-compliance.[12] The Supreme Court of New South Wales Practice Note SC Gen 23 imposes verification requirements and disclosure obligations on AI-assisted submissions.[13] The duty is the same across borders. Only the local labels differ.

Concern Resolved:

You will not lose your license for using AI. You may lose it for filing AI-generated content you did not verify. The duties are not new. They are existing duties applied to a new tool.

Part Three

“Will my client’s data leak?”

The Samsung Lesson

In March 2023, engineers at Samsung Electronics inadvertently exposed proprietary semiconductor source code, internal meeting notes, and confidential test sequences by pasting them into ChatGPT for help debugging and summarizing. Within weeks, three separate incidents were documented. Samsung had no ability to retrieve the data, which had passed into OpenAI’s systems. The company eventually imposed a strict upload limit and, for a time, banned generative AI use entirely.[14] Lawyers should treat the Samsung incident as a parable. The engineers were not malicious. They were trying to do their jobs faster. The data left the building anyway.

Three Mechanisms of Leak

Client data can leave a lawyer’s control through generative AI in three distinct ways, each with a different mitigation.

- **Training on inputs.** Many consumer AI services, by default, use submitted prompts to improve their models. The data may surface in future outputs to other users. Enterprise tiers typically disable this, but only if the contract and the configuration both confirm it.
- **The sub-processor chain.** A legal-specific AI tool may be a thin wrapper over a foundation model hosted by another vendor, which in turn runs on a cloud provider in a jurisdiction the lawyer never selected. Each link is a potential disclosure point. Each requires its own data processing terms.
- **Breach and inadvertent access.** AI vendors are software companies. They suffer breaches. They have employees with access to logs. They retain data longer than the lawyer expects unless the contract says otherwise.

Privilege and Waiver

Whether transmitting privileged information to a third-party AI vendor waives privilege is unsettled and varies by jurisdiction. In most U.S. jurisdictions the analysis tracks the existing law on cloud service providers, e-discovery vendors, and outside contractors: disclosure to a vendor that is bound by confidentiality and acts as the lawyer’s agent generally does not waive privilege. The argument breaks down when the vendor’s terms permit training on the data, when retention exceeds what the engagement requires, or when the data crosses borders into a regime that does not recognize the privilege at all.[15] The conservative position, and the one supported by ABA Formal Opinion 512 and the Florida Bar’s Advisory Opinion 24-1, is to obtain informed client consent before submitting confidential information to a third-party generative AI tool, and to refuse submission to any tool whose terms or configuration permit training on inputs.[16]

What Confidentiality Now Requires

ABA Model Rule 1.6(c) requires lawyers to make reasonable efforts to prevent inadvertent or unauthorized disclosure of information relating to the representation. The standard is fact-specific and tracks evolving technological norms.[17] In the AI context, the floor now includes:

- Reading the data processing terms before approving any AI tool for client work, and confirming whether the vendor trains on inputs, retains prompts, and permits human review.

- Maintaining an approved-tool list that distinguishes between consumer-grade tools (prohibited for client data) and enterprise-tier tools that contractually disable training and limit retention.
- Obtaining informed consent before placing any confidential client information into a third-party generative AI tool, with consent that is specific to AI use rather than buried in a generic technology clause.
- Treating cross-border data flows as a separate analysis. The Italian Garante's 2023 temporary ban on ChatGPT, and its subsequent fifteen-million-euro fine in December 2024, demonstrated that AI processing in one jurisdiction can produce regulatory exposure in another.[18]

Concern Resolved:

Your client's data does not have to leak. It will leak only if you submit it to a tool that trains on inputs, retains prompts longer than you authorize, or stores data in a jurisdiction your engagement does not contemplate. The contract and the configuration are the controls.

Part Four

“Can I trust the output?”

Hallucinations in One Paragraph

A large language model generates text by predicting the most probable next token. It does not retrieve verified information. It does not know what is true. When asked for a case to support a proposition, it produces output that fits the statistical pattern of a citation, regardless of whether the case exists. The Stanford HAI study by Magesh and colleagues, published in the Journal of Empirical Legal Studies, found that purpose-built legal AI tools using retrieval-augmented generation still hallucinate on roughly 17 to 33 percent of legal research queries. General-purpose models hallucinate on substantially more.[19] The companion guide in this series, Managing AI Hallucinations, addresses the mechanism and the verification protocol in depth.

Bias and Disparate Impact

Hallucinations are not the only output risk. AI models trained on historical data reproduce historical biases. In legal contexts this matters most where AI is used in adverse decisions: hiring, lending, housing, immigration, sentencing, parole risk assessment. The SRA’s Risk Outlook flagged bias amplification as one of the central risks in the legal market.[10] The Law Society of Ontario white paper makes the same point, explicitly listing discrimination and harassment among the professional obligations affected by AI use.[11] Lawyers advising clients on AI deployment, or using AI to make recommendations that affect clients, should treat bias as a foreseeable harm, not as an exotic edge case.

Explainability and the Black Box

Foundation models do not explain their reasoning. A retrieval-augmented system can show the documents it retrieved, but cannot explain why it weighed them as it did. For most internal uses this opacity is tolerable. For decisions that affect a client’s rights, it is not. Where the law requires reasons, the lawyer must be able to provide reasons that do not depend on the model’s assertion of its own correctness.

The Verification Standard

ABA Formal Opinion 512 establishes that the verification required is task-specific and depends on what the AI tool is being used for. Document review and idea generation require less independent verification than legal research and citation.[9] The Singapore Registrar’s Circular and the NSW Practice Note both impose explicit verification requirements: every citation, every quotation, every authority must be confirmed by the lawyer before the work product reaches a tribunal.[12][13] The operating principle is the one stated in the companion guide: an AI output is a draft, not a deliverable. It carries no legal consequence until a human verifies it.

Concern Resolved:

You can trust AI output the same way you trust an associate’s first draft: not at all, until you have read it, checked the sources, and applied your own judgment. Verification is the trust mechanism. There is no other.

Part Five

“Who pays when it goes wrong?”

Malpractice Exposure

The malpractice analysis is straightforward and uncomfortable. The standard of care is what a reasonably competent lawyer would do. As ABA Formal Opinion 512 and the cluster of state bar guidance make clear, that standard now includes a reasonable understanding of any AI tool the lawyer uses, plus verification of its output before reliance.[9] A lawyer who files AI-generated content without verification has, in most jurisdictions, fallen below the standard of care. Damages flow from ordinary negligence principles. The plaintiff’s expert will not have to work hard.

The standard runs in both directions. Some commentators argue that the failure to use available AI tools, where doing so would meaningfully improve cost or accuracy, may itself become a competence concern under Rule 1.1’s duty to keep abreast of the benefits and risks of relevant technology.[20] That argument is not yet established by case law, but the direction of travel is unmistakable.

The Insurance Coverage Gap

Most professional liability policies do not explicitly address AI-related errors. Whether an AI-caused malpractice claim falls within coverage depends on how the policy defines “professional services” and whether the insurer treats AI-assisted work as falling within that definition. The ABA Journal’s February-March 2025 review of professional liability coverage for AI mistakes concluded that significant uncertainty exists across standard policies.[21] The risk runs in both directions. If a lawyer cannot demonstrate reasonable care in using AI tools, an insurer may argue that no covered “professional service” occurred because the lawyer delegated judgment to a machine without adequate oversight.

Specialized AI insurance products are emerging but remain immature. Munich Re’s aiSure has been writing AI performance insurance since 2018 and has expanded to cover hallucinations, intellectual property infringement, and systemic discrimination. Armilla, backed by Chaucer and Axis Capital, launched a purpose-built AI liability product in 2025 that contemplates hallucinations, model drift, and deviations from expected behavior.[22] These products are oriented to enterprise AI deployers and AI vendors, not directly to lawyers, but they are signals of where the coverage market is moving.

The Vendor Contract Is the Risk Allocation

When AI causes harm, three contracts decide who pays: the firm’s engagement letter with the client, the firm’s professional liability policy, and the firm’s contract with the AI vendor. The third is the one most often overlooked. Liability caps, indemnification scope, accuracy warranties, vendor insurance requirements, and audit rights in the AI vendor agreement determine whether the firm bears the cost of an AI failure or can pass it to the vendor. Standard vendor terms are written for the vendor, not the firm. They typically cap liability at fees paid, exclude consequential damages, and disclaim accuracy warranties entirely. A firm that signs without negotiation has implicitly accepted the entire risk.

Five Questions for Any AI Vendor

Before signing or renewing any AI vendor agreement that will touch client matters, get written answers to the following:

- Will you use our prompts, outputs, or any other Customer Data to train, fine-tune, or improve any model, whether yours or a third party's? If yes, can that be turned off contractually and confirmed in our tenant configuration?
- How long do you retain prompts and outputs by default, and what is the shortest retention period you will commit to in writing?
- Who are your sub-processors, including the foundation model provider, the cloud host, and any third-party services? What is your notice period before you change one?
- What are the indemnification scope and the liability cap, and do they survive your standard liability cap for IP infringement and data breach?
- What jurisdictions store and process our data, and what cross-border transfer mechanisms do you rely on for EU, UK, and other regulated personal data?

Model Vendor Contract Clauses

The clauses below are illustrative starting points for negotiating an AI vendor agreement. They should be adapted to the firm's standard contracting framework, the specific Service, and the regulatory regime that applies to the data involved. They are not a substitute for review by the lawyer who will be accountable for the relationship.

Model Vendor Contract Clauses

1. No Training Use of Customer Data.

Vendor shall not, and shall ensure that no sub-processor shall, use Customer Data, including prompts, outputs, embeddings, or metadata, for the training, fine-tuning, evaluation, or improvement of any artificial intelligence model, whether Vendor's own or any third party's. Customer Data shall be used solely to provide the Service to Customer in accordance with this Agreement.

2. Retention and Deletion.

Vendor shall not retain Customer Data for longer than thirty (30) days from the date of submission, except as strictly required to provide the Service or as required by applicable law. Upon Customer's written request, Vendor shall delete identified Customer Data within seven (7) business days and shall certify deletion in writing. On termination or expiration of this Agreement, Vendor shall delete all Customer Data within thirty (30) days and shall certify deletion.

3. Sub-Processor Disclosure and Change Control.

Vendor shall maintain a current list of all sub-processors that may process Customer Data, including the foundation model provider, the hosting provider, and any third-party services involved in delivering the Service, and shall make that list available to Customer. Vendor shall provide Customer at least thirty (30) days' prior written notice before adding or replacing any sub-processor. Customer may terminate this Agreement, with a pro rata refund of any prepaid fees, if Customer reasonably objects to the change.

4. Accuracy Disclaimer and Output Indemnification.

Customer acknowledges that the Service may produce outputs that are inaccurate, incomplete, or fabricated, and that Customer is responsible for verification before reliance. Notwithstanding the foregoing, Vendor shall indemnify Customer against third-party claims alleging that the Service's outputs, when used in accordance with the Documentation, infringe the intellectual property rights of any third party. The aggregate liability cap in Section [X] shall not apply to Vendor's indemnification obligations under this clause or to Vendor's breach of its data protection obligations.

5. Audit and Information Rights.

On reasonable prior notice and not more than once per calendar year, Customer or its designated auditor may review Vendor's SOC 2 Type II report, ISO 27001 certification, and any other security and privacy documentation reasonably necessary for Customer to demonstrate its own compliance with applicable professional, regulatory, and contractual obligations relating to Customer Data.

Three Questions for Your Broker

Before the next policy renewal, contact your broker and ask three questions:

- Does the current policy cover malpractice claims arising from reliance on AI-generated output, including hallucinations and biased recommendations?

- Does the policy distinguish between AI-assisted work and AI-generated work, and if so, what documentation is required to demonstrate that the work qualifies as AI-assisted?
- If the current policy does not provide affirmative AI coverage, what supplemental coverage is available, and at what premium?

Practical Takeaway:

Three contracts allocate the cost of AI failure: the engagement letter, the malpractice policy, and the vendor agreement. If you have not read all three with AI risk specifically in mind, you do not know who pays.

Concern Resolved:

Liability is allocable. It is not, however, automatic. The firms that allocate it deliberately, in writing, before an incident, are the firms that recover when one happens.

Part Six

“Do I have to tell my client and my court?”

Client Disclosure and Informed Consent

ABA Formal Opinion 512 takes a clear position: where AI use will involve disclosure of confidential client information, the lawyer should obtain the client’s informed consent before doing so, and boilerplate consent in engagement letters is not sufficient. The opinion further holds that the duty of communication under Model Rule 1.4 may require disclosure of AI use even where confidential information is not involved, when the use is material to the representation or to the client’s decisions about the matter.[9] The Florida Bar reached the same conclusion in Advisory Opinion 24-1: informed consent is recommended before utilizing a third-party generative AI program if the use involves disclosure of confidential information.[16]

Engagement Letter Language

Generic technology consent language does not satisfy the informed consent standard ABA Opinion 512 describes. Effective AI disclosure language identifies which AI tools the firm uses, the categories of tasks for which they are used (research, drafting, document review, summarization), the verification protocols the firm applies before AI-assisted work product reaches the client or a tribunal, and the known limitations of the tools, including the residual risk of error. Specificity is what converts a boilerplate clause into informed consent.

> **Model Engagement Letter Clause: Use of Artificial Intelligence**
>
> Use of Artificial Intelligence Tools. The Firm uses generative artificial intelligence tools to assist with legal research, document drafting, document review, summarization, and similar tasks. The Firm does not delegate professional judgment or final work product to these tools. Each AI-assisted output is reviewed and verified by a lawyer before it is relied upon, sent to the Client, or filed with any tribunal. The Firm uses only enterprise-tier tools whose terms prohibit the vendor from using the Client’s information to train its models, and the Firm maintains contractual data-protection commitments with each AI vendor it uses on the Client’s matter. The Firm will not submit information that the Client has designated as highly confidential to any AI tool without the Client’s prior written consent. Time saved through the use of AI tools will not be billed to the Client. The actual subscription cost of AI tools used on the Client’s matter is treated as Firm overhead and is not separately charged to the Client unless the parties agree in writing in advance. The Client may at any time instruct the Firm not to use AI tools on a particular matter or task.

Fees and Billing for AI Use

Three positions emerge from the major guidance. First, lawyers may not charge clients for time spent learning a generative AI technology that becomes part of the firm’s general competence, just as they cannot charge for time spent learning Westlaw.[9] Second, where AI reduces the time required to perform a task, lawyers may not bill the client for time saved.[16] The California State Bar’s practical guidance is direct: a lawyer must not charge for the time saved by using GenAI.[23] Third, lawyers may charge clients for the actual cost of AI tools used in the matter, provided this is disclosed in advance and the client consents. Whether AI cost is best treated as overhead or as a per-matter charge is a business decision, but the disclosure obligation is not.

Court Disclosure and Standing Orders

Court disclosure is now mandatory in a growing number of jurisdictions. Judge Brantley Starr of the Northern District of Texas requires every attorney appearing before him to certify either that no portion of any filing was drafted by generative AI, or that any AI-drafted language was checked for accuracy by a human being.[24] The Supreme Court of New South Wales Practice Note SC Gen 23, effective February 2025, requires affidavits, witness statements, and character references to disclose non-use of generative AI, and requires verification statements for citations and authorities in submissions and skeletons of argument.[13] The Singapore Registrar's Circular No. 1 of 2024 places full responsibility for AI output on the user across the Supreme Court, State Courts, and Family Justice Courts, with explicit cost and disciplinary consequences.[12]

The practical posture is to assume disclosure may be required and to maintain the documentation that permits accurate disclosure if asked. A firm that cannot answer the question "which tools touched this filing" cannot make a truthful certification when one is requested.

Concern Resolved:

Yes, in many circumstances you must tell your client, and in a growing list of courts you must tell the tribunal. The disclosure is not a confession of weakness. It is the documentation that protects you when something goes wrong.

Part Seven

“What if someone on my team misuses it?”

AI as Nonlawyer Assistant

ABA Formal Opinion 512 took an analytic step that simplifies the supervision question considerably: it treats generative AI tools as nonlawyer assistants for purposes of the supervisory rules. The obligations that apply when a paralegal or contract attorney does work apply equally when an AI tool does work. The supervising lawyer is responsible for ensuring the work product is competent and consistent with professional obligations.[9] The Florida Bar’s Opinion 24-1 reached the same framing.[16] The Law Society of Ontario’s white paper applied the same analysis under the Ontario rules.[11]

Rules 5.1 and 5.3 in Practice

ABA Model Rule 5.1 imposes managerial and supervisory responsibility on partners and lawyers with managerial authority for the conduct of subordinate lawyers. Rule 5.3 extends the same framework to nonlawyer assistants. Together they require that firms establish reasonable policies and supervision to ensure that AI use within the firm meets the lawyers’ professional obligations. In practice this means the supervising lawyer must understand what AI tools are in use, what tasks they are being used for, what verification protocols apply, and what to do when an associate’s AI-assisted work product is incorrect.

Agentic Systems and the Supervision Gap

The supervision question becomes more complicated as AI systems take more autonomous action. An agentic system that drafts, files, or sends without a human approval step in the loop is indistinguishable, from the regulator’s perspective, from a paralegal who acts without supervision. The supervision must be designed in, not added on. Where the workflow does not preserve a human checkpoint at the moment work product becomes consequential, the supervisory architecture has failed regardless of how careful the operator believes themselves to be.

The Policy and Training Floor

Most published guidance now treats a written AI policy and mandatory training as the floor for compliance with the supervisory rules. The policy should identify approved tools, prohibited use cases, required verification protocols, the categories of data permitted in each tool, and the consequences for noncompliance. Training should cover how the tools work, what hallucinations are, how to verify outputs, what the rules require, and what has gone wrong in the published cases. The ABA opinion, the Florida Opinion 24-1, the California Practical Guidance, and the LSO white paper all point to this same floor.[9162311]

Model AI Use Policy: Core Provisions

The provisions below are a starting point for a firm AI use policy. They should be tailored to the firm's tools, practice areas, and regulatory exposures, and adopted by the partnership or governance body responsible for professional standards.

1. Scope.

This policy applies to every lawyer, paralegal, secretary, contractor, and other person who uses any artificial intelligence tool in connection with any matter for any client of the Firm.

2. Approved Tools.

Only tools listed on the Firm's Approved AI Tools List, in the configuration the Firm has provisioned, may be used for any work touching client matters. Use of consumer or personal-account versions of any tool, including consumer versions of approved tools, is prohibited for client matters.

3. Data Classification.

Client confidential information may be submitted only to Approved Tools designated for client work. Personal data subject to GDPR, the UK GDPR, or other privacy regimes may be submitted only to Approved Tools that the Firm has confirmed satisfy the applicable cross-border transfer requirements. Information designated by the client as highly confidential may not be submitted to any AI tool without the client's prior written consent recorded on the matter file.

4. Verification.

Every citation, quotation, and material factual assertion in any AI-assisted work product must be verified against an authoritative source by a lawyer before the work product is sent to a client, served on opposing counsel, or filed with any tribunal. The lawyer who signs or sends the work product is responsible for verification, regardless of which person generated the AI-assisted draft.

5. Disclosure.

AI use shall be disclosed to clients consistent with the Firm's engagement letter template, and to any tribunal where required by court order, practice direction, or local rule. Lawyers are responsible for monitoring the AI disclosure requirements of every court in which they appear and for documenting compliance in the matter file.

6. Incident Reporting.

Any AI-generated error that reaches a client or tribunal, or that is caught only after the work product was relied upon internally, must be reported to the AI Governance Committee within five business days. Reports are entered in the AI Incident Log and inform tool decisions, training updates, and policy revisions.

7. Training and Acknowledgment.

Every person within scope of this policy shall complete the Firm's annual AI training and shall acknowledge this policy in writing. Use of any AI tool on client matters by a person who has not completed current training is prohibited.

8. Enforcement.

Violations of this policy will be addressed under the Firm's standard disciplinary procedures and may result in additional training, revocation of access to AI tools, or other employment action proportionate to the violation.

Concern Resolved:

If someone on your team misuses AI, the rules will hold you responsible the same way they would if they misused any other tool the firm provides. The cure is the same: a written policy, mandatory training, an approved tool list, and a documented verification protocol that the team actually follows.

Part Eight

“Which rules actually apply to me?”

The Multi-Jurisdictional Reality

Most lawyers practice across multiple regulatory regimes whether they acknowledge it or not. A New York firm representing a German client whose data sits on a U.S. cloud, with associates working remotely from a third state, faces the rules of all of them. AI compounds the complexity because the model, the vendor, the cloud, and the user can each be in a different jurisdiction. Three observations help organize the field.

- First, the duties converge. Competence, confidentiality, supervision, candor, communication, and reasonable fees appear in nearly every professional regime examined for this guide.
- Second, the disclosure obligations diverge. Whether and when to disclose AI use to clients and tribunals varies sharply across jurisdictions and is changing fastest.
- Third, the data law layer is independent. GDPR, CCPA, sectoral privacy regimes, and regulations like the EU AI Act apply on top of professional rules and may produce stricter obligations than the bar imposes.

United States: ABA and the State Bars

ABA Formal Opinion 512 is the touchstone, but the binding obligations come from each state bar. The Florida Bar’s Advisory Opinion 24-1 (January 2024) and the California State Bar’s Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law (November 2023) are the most extensively developed.[16][23] New York, Texas, Pennsylvania, and additional jurisdictions have issued opinions and standing orders of varying scope. Where a federal court has issued a standing order, as Judge Starr did in the Northern District of Texas, that order controls regardless of the practitioner’s home bar.[24]

United Kingdom, Canada, Australia, Singapore

The Solicitors Regulation Authority’s Risk Outlook on AI in the legal market sets the regulatory frame for England and Wales, emphasizing accuracy, accountability, and the duty to understand the tools in use. The SRA has been explicit that consumers face risk from lawyers who do not understand the technology they implement.[10] The Law Society of Ontario’s April 2024 white paper, Licensee Use of Generative Artificial Intelligence, sets out professional obligations under the Ontario rules and is accompanied by quick-start checklists and best-practice tips through the LSO Technology Resource Centre. Other Canadian provinces, including Alberta and British Columbia, have issued parallel guidance.[11] In Australia, the Supreme Court of New South Wales Practice Note SC Gen 23, effective February 3, 2025, applies to civil and commercial proceedings before that court and sets out detailed disclosure and verification requirements.[13] The Federal Court of Australia has not yet issued a formal practice note but expects responsible use consistent with existing obligations. In Singapore, the Registrar’s Circular No. 1 of 2024 applies to all matters in the Supreme Court, the State Courts, and the Family Justice Courts and places full responsibility for AI output on the user.[12]

European Union: GDPR and the AI Act

European lawyers and any practitioner whose work touches data of EU subjects must reckon with two regimes that operate independently of professional conduct rules. The General Data Protection Regulation governs the processing of personal data, including data submitted to AI tools. The Italian Garante's March 2023 emergency decision halting ChatGPT processing of Italian data subjects, and its subsequent fifteen-million-euro fine in December 2024, demonstrated that enforcement is real and not limited to the largest providers.[18] The Artificial Intelligence Act, Regulation (EU) 2024/1689, entered into force on August 1, 2024, with the bulk of obligations applying from August 2, 2026. The Act classifies certain legal-services applications as high-risk and imposes risk-management, documentation, human oversight, and post-market monitoring obligations on providers of those systems.[25] Lawyers using high-risk AI in EU practice should treat the August 2026 deadline as imminent.

Cross-Border Note:

When the model, the vendor, the cloud, and the user are in different jurisdictions, you do not get to choose which rules apply. They all do. Identify each layer of your AI stack and map each to its regulatory regime before you place client data into it.

Part Nine

“How do I make this defensible?”

The Concerns-to-Controls Map

Every concern this guide addresses has a corresponding control. The map below summarizes the eight concerns and the control that addresses each. The body of the guide explains the mechanism behind each. The map is the worklist.

Concern	Control
Sanctions and license loss	Verification protocol on every AI-assisted filing; documented Rule 11 inquiry
Client data leak	Approved-tool list; informed consent; ban on consumer tools for client matters
Untrustworthy output	Five-step verification; treat output as draft, not deliverable
Liability when it goes wrong	Engagement letter, malpractice policy, and vendor contract reviewed for AI risk
Disclosure to client and court	Specific AI consent in engagement letter; audit-trail documentation
Misuse by team	Written AI policy; mandatory training; supervision under Rules 5.1 and 5.3
Multi-jurisdictional rules	Map each layer of the AI stack to its regulatory regime; conservative posture across regimes
Defensibility	Governance committee; incident log; annual reassessment

Governance, Tool List, Intake Triage

Three governance artifacts make the controls operational. An AI governance committee with representation from at least one equity partner, IT or information security, and the major practice groups, meeting at least quarterly, with authority over tool selection, policy, and incident review. An approved-tool list that distinguishes consumer-grade and enterprise tools, records the data processing terms negotiated for each, and identifies the categories of data permitted in each. An intake triage process that classifies new use cases into the Red, Yellow, and Green tiers described in the companion hallucinations guide, with corresponding verification requirements.

Training, Incident Log, Annual Reassessment

Training closes the gap between policy and practice. Every lawyer and staff member who uses AI on client matters should complete training that covers what hallucinations are, how to verify, what the rules require, and what has gone wrong in published cases such as Mata, Park, Noland, and Zhang. The training is mandatory and refreshed annually. An incident log records every instance in which an AI-generated error was caught during verification or reached a client or tribunal. The log informs tool decisions, policy revisions, and the firm's defense if its governance practices are ever examined. Annual reassessment of the approved tool list, the policy, the training curriculum, and the incident log keeps the program aligned with rapidly evolving guidance.

Maturity Model and Self-Audit

Most firms can place their current AI practice on a four-stage spectrum.

- **Reactive.** No written policy. AI use is informal and ad hoc. Verification depends on individual discipline. The firm is exposed.
- **Compliant.** Written policy, approved tool list, and training in place. Verification is required. The firm meets the floor that ABA Opinion 512 and the leading state bars describe.
- **Proactive.** Governance committee, incident log, vendor contract review, and explicit engagement letter language. The firm is positioned to defend its practices if questioned.
- **Strategic.** AI is integrated into matter intake and pricing, supervised through deliberate human checkpoints, and treated as a competitive advantage rather than a compliance burden.

Move one stage at a time. The firms that survive the next five years of AI in legal practice will be the ones whose maturity advances faster than the regulatory landscape changes.

Concern Resolved:

Defensibility is not perfection. It is documentation. A firm that can show its policy, its training log, its approved-tool list, its verification protocol, and its incident log has a defense. A firm that cannot does not.

Part Ten

What Comes Next

AI tools will continue to improve. Hallucination rates will fall. Vendor terms will mature. Insurance coverage will become more affirmative. Bar guidance will continue to converge. None of this changes the fundamental posture lawyers should adopt now.

The concerns are real. Each has a control. The control is documented in policy, embedded in workflow, and supported by training. The lawyers and firms that build that infrastructure now will use AI responsibly across the next decade. Those who do not will keep finding themselves in the Charlotin database for reasons they could have prevented.

The standard is not perfection. It is diligence. Verify. Document. Disclose. Train. Reassess. These are not new obligations. They are existing obligations applied to a new tool. The tool is powerful. The obligations have not changed.

Glossary of Key Terms

Algorithmic Bias.
The tendency of an AI system to produce outputs that systematically advantage or disadvantage particular groups, typically as a result of biases present in training data or model design. Bias can produce professional responsibility exposure when AI-assisted work affects adverse decisions about clients or third parties.

Approved-Tool List.
A firm-maintained inventory of AI tools approved for use, classified by data sensitivity, with documented data processing terms, retention policies, and permitted use cases. Central to defensibility under the supervisory rules.

Confidentiality (Model Rule 1.6).
The lawyer's duty to protect information relating to the representation of a client. Rule 1.6(c) requires reasonable efforts to prevent inadvertent or unauthorized disclosure, a standard that scales with the sensitivity of the information and prevailing technological norms.

Data Residency.
The geographic location where data is stored or processed. Material to AI use because models, vendors, and cloud infrastructure may sit in jurisdictions with different privacy and disclosure regimes than the client or matter.

EU AI Act (Regulation (EU) 2024/1689).
European Union regulation establishing risk-based obligations for AI systems, in force from August 1, 2024, with most obligations applying from August 2, 2026. Certain legal-services applications fall into the high-risk category and trigger documentation, oversight, and monitoring obligations.

Foundation Model.
A large AI model trained on broad data and adaptable to many downstream tasks. Examples include GPT, Claude, Gemini, and Llama. Legal-specific tools typically build on top of foundation models with additional retrieval or fine-tuning layers.

Hallucination.
An AI output that is factually incorrect, fabricated, or unsupported by source material, presented with apparent confidence. The leading source of legal sanctions arising from AI use.

Informed Consent.
Client agreement obtained after disclosure sufficient to permit a reasonable client to evaluate the decision. Generic technology consent in engagement letters is not sufficient under ABA Formal Opinion 512 to support disclosure of confidential client information to a third-party AI tool.

Privilege.
The protection that shields communications between lawyer and client (and, in many regimes, the lawyer's work product) from compelled disclosure. AI processing creates a third-party touch that requires careful structuring to preserve.

Reasonable Lawyer Standard.
The objective benchmark used in most jurisdictions to evaluate professional conduct: what would a competent practitioner have done under the same circumstances. The standard governs AI use even where no AI-specific rule has been adopted.

Retrieval-Augmented Generation (RAG).
An AI architecture that retrieves relevant documents from a knowledge base before generating a response, intended to ground outputs in source material. Reduces hallucination rates but does not eliminate them.

Sub-Processor.
A third party that processes data on behalf of a primary AI vendor. The chain of sub-processors is the actual surface area of data exposure and should be enumerated in the vendor agreement.

Supervision (Model Rules 5.1 and 5.3).
The duties imposed on partners and lawyers with managerial authority to ensure that subordinate lawyers and nonlawyer assistants conform to professional rules. ABA Formal Opinion 512 treats AI tools as nonlawyer assistants for this purpose.

Training on Inputs.
The practice by which an AI vendor uses user-submitted prompts to further train its models. Standard in consumer tiers; typically disabled in enterprise tiers, but only if both the contract and the configuration confirm it.

Verification Protocol.
A documented, repeatable procedure for confirming the accuracy of AI-generated output before reliance. Required as a matter of professional responsibility in nearly every jurisdiction examined.

Endnotes

1. ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512, “Generative Artificial Intelligence Tools” (July 29, 2024); Florida Bar Board of Governors, Ethics Opinion 24-1, “Lawyers’ Use of Generative Artificial Intelligence” (January 19, 2024); California State Bar, “Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law” (November 16, 2023); Solicitors Regulation Authority, “Risk Outlook Report: The Use of Artificial Intelligence in the Legal Market” (2024); Law Society of Ontario, “Licensee Use of Generative Artificial Intelligence” (April 2024); Supreme Court of New South Wales, Practice Note SC Gen 23 (issued November 21, 2024; effective February 3, 2025); Singapore Registrar’s Circular No. 1 of 2024, “Guide on the Use of Generative Artificial Intelligence Tools by Court Users”; Regulation (EU) 2024/1689 (Artificial Intelligence Act).

2. See, e.g., Gunder, “Rule 11 Is No Match for Generative AI,” 27 Stan. Tech. L. Rev. 308 (Spring 2024), analyzing the objective reasonableness standard as applied to AI-generated court filings.

3. Federal Rule of Civil Procedure 11(b)(2)–(3) (requiring legal contentions to be warranted by existing law and factual contentions to have evidentiary support).

4. Damien Charlotin, “AI Hallucination Cases” database, available at damiencharlotin.com/hallucinations. As of early 2026 the database contained more than 1,300 documented decisions across multiple jurisdictions.

5. Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. 2023). $5,000 sanction imposed for filing six fabricated case citations generated by ChatGPT and providing fabricated opinion texts when challenged. Court found “subjective bad faith” in the failure to verify.

6. Park v. Kim, 91 F.4th 610 (2d Cir. 2024). Second Circuit referred attorney to its Grievance Panel for citing a nonexistent case generated by ChatGPT and making no inquiry into its validity.

7. Noland v. Land of the Free, L.P., Cal. Ct. App., 2d Dist. (Sept. 2025). $10,000 sanction imposed; counsel reported to the State Bar of California. First published California Court of Appeal opinion addressing AI-fabricated legal authority.

8. Zhang v. Chen, 2024 BCSC 285 (Supreme Court of British Columbia). Court ordered counsel to personally compensate opposing counsel for time spent unwinding ChatGPT-generated fabricated authority. Law Society of British Columbia opened investigation.

9. ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512, “Generative Artificial Intelligence Tools” (July 29, 2024). Addresses competence (Rule 1.1), confidentiality (Rule 1.6), communication (Rule 1.4), candor (Rules 3.1 and 3.3), supervisory responsibilities (Rules 5.1 and 5.3), and reasonable fees (Rule 1.5).

10. Solicitors Regulation Authority, “Risk Outlook Report: The Use of Artificial Intelligence in the Legal Market” (2024), and related SRA statements that consumers face risk from lawyers who do not understand the technology they implement.

11. Law Society of Ontario, “Licensee Use of Generative Artificial Intelligence” (white paper, April 2024), accompanied by quick-start checklist and best-practice tips through the LSO Technology Resource Centre. Identifies competence, confidentiality, supervision, licensee-client relationships, fees and disbursements, and discrimination as professional obligations engaged by AI use.

12. Singapore Registrar’s Circular No. 1 of 2024, “Guide on the Use of Generative Artificial Intelligence Tools by Court Users.” Applies to all matters in the Supreme Court, the State Courts, and the Family Justice Courts. Non-compliance may result in adverse cost orders, evidentiary discounts, or disciplinary action.

13. Supreme Court of New South Wales, Practice Note SC Gen 23, “Use of Generative Artificial Intelligence” (issued November 21, 2024; effective February 3, 2025). Requires disclosure statements for affidavits, witness statements, and character references and verification of citations in submissions.

14. See, e.g., “Incident 768: ChatGPT Implicated in Samsung Data Leak of Source Code and Meeting Notes,” AI Incident Database; Samsung Electronics internal notice to employees, April 2023; company-wide restrictions on generative AI use following the incident.

15. See ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 477R, “Securing Communication of Protected Client Information” (May 22, 2017), updated guidance on the duty to protect client data in electronic communications, and the cluster of state bar opinions addressing cloud services that informs the privilege analysis for AI vendor processing.

16. The Florida Bar Board of Governors, Ethics Opinion 24-1, “Lawyers’ Use of Generative Artificial Intelligence” (January 19, 2024). Recommends that lawyers obtain affected clients’ informed consent before using third-party generative AI programs that involve disclosure of confidential information.

17. ABA Model Rule 1.6(c) (“A lawyer shall make reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client.”). See also Comment 18 to Rule 1.6.

18. Italian Garante per la protezione dei dati personali, emergency decision halting ChatGPT processing of Italian data subjects (March 30, 2023); subsequent administrative fine of €15 million imposed on OpenAI (December 2024) for GDPR violations relating to training-data processing and lack of an adequate legal basis.

19. Magesh, Surani, Dahl, Suzgun, Manning, and Ho, “Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools,” 22 J. Empirical Legal Stud. 216 (2025). Found hallucination rates of approximately 17% (Lexis+ AI), 33% (Westlaw AI-Assisted Research), and 43% (GPT-4) across federal case-law queries.

20. See discussion of evolving competence obligations in commentary on ABA Formal Opinion 512 and Comment 8 to Model Rule 1.1, addressing the duty to keep abreast of the benefits and risks of relevant technology.

21. ABA Journal, “Does Your Professional Liability Insurance Cover AI Mistakes? Don’t Be So Sure” (February-March 2025). Identifies significant coverage uncertainty for AI-related malpractice claims under standard professional liability policies.

22. Munich Re, aiSure product line, expanded to address hallucinations, intellectual property infringement, and systemic discrimination; Armilla Insurance, purpose-built AI liability policy launched 2025, backed by Chaucer and Axis Capital, contemplating coverage for hallucinations, model degradation, and deviations from expected behavior.

23. California State Bar, “Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law” (November 16, 2023). A lawyer must not charge the client for the time saved by using generative AI; lawyers may charge for the time actually spent on the work product.

24. Standing Order, Judge Brantley Starr, U.S. District Court for the Northern District of Texas (May 30, 2023). Mandatory Certification Regarding Generative Artificial Intelligence requires attorneys to certify either non-use of generative AI in the filing or human verification of any AI-generated language.

25. Regulation (EU) 2024/1689 (Artificial Intelligence Act), entered into force August 1, 2024; majority of obligations applicable from August 2, 2026. High-risk classification under Annex III applies to specified uses including those affecting access to justice and the administration of law.