THE 2025 AI LIABILITY RISK ASSESSMENT MATRIX

For Legal Professionals & Compliance Officers
Version 2.0 (Effective Jan 2025)

1. EXECUTIVE SUMMARY & PURPOSE

The era of treating all AI output as equally risky is over. A "zero-tolerance" policy stifles innovation, while a "laissez-faire" policy invites malpractice.

This matrix provides a tiered framework for AI adoption. It categorizes legal tasks by **Liability Risk**—the probability and severity of professional negligence claims, reputational damage, or court sanctions resulting from AI error.

Core Principle: The level of human verification must rise in direct proportion to the **finality** of the output and the **reliance** placed upon it by third parties (courts, clients, opposing counsel).

2. THE RISK ASSESSMENT MATRIX

Risk Level	**Task Category**	**Specific Liability Vector**	**Approved Tooling Class**	**Mandatory Verification Protocol**
LEVEL 1: LOW *(Internal / Ideation)*	• Brainstorming legal arguments • Summarizing long transcripts • Internal memos (Draft 1) • Translation (Gist only)	**Low.** Errors here are "process errors," not "outcome errors." Unlikely to leave the firm or damage a client's case if caught later.	• Public LLMs (ChatGPT, Claude) • Enterprise Wrappers	**Protocol A: Logic Check** Review for coherence. No citation verification required at this stage.
LEVEL 2:	• Client status	**Moderate.**	• Enterprise	**Protocol B:**

MEDIUM *(Client Communication)*	updates • Drafting routine NDAs • Clause generation (Standard) • Marketing content	Risk of misstating a deadline or promising an outcome. "Silent Cyber" risk if confidential data is leaked to public models.	Wrappers (Data Privacy Mode ON) • Secure RAG Tools	**Substantive Review** Lawyer must read every word. Verify specific dates and promises against the case file.
LEVEL 3: HIGH *(Substantive Legal Work)*	• Drafting Case Briefs • Contract Analysis (Risk finding) • Statutory Interpretation • Researching Precedent	**High.** Risk of **Hallucination** (Fake Law) or **Sycophancy** (Bias confirmation). Errors here directly impact legal strategy and malpractice liability.	• **Legal-Grade RAG Only** (Harvey, Lexis+, CoCounsel, Thomson Reuters) • *NO Public LLMs*	**Protocol C: Source Trace** Every claim must be clicked through to the primary source. "Grounding" is not truth; verify the source exists.
LEVEL 4: CRITICAL *(Court / Final Opinion)*	• Court Filings (Motions/Pleadings) • Final Legal Opinions • Citations of Authority • Settlement Calculations	**Severe. Sanctions Territory.** Submitting fake citations or hallucinated facts constitutes "Misleading the Court" (See *Mata*, *Choksi*).	• **None for Final Polish.** AI may draft, but a Human must finalize independent of the tool.	**Protocol D: Forensic Audit** Independent verification of every citation in an outside database (e.g., Westlaw/Lexis standard search) *not* using the AI tool.

3. PROTOCOL DEFINITIONS

Protocol A: Logic Check (The "Sniff Test")

- **Action:** Read the output to ensure it flows logically and addresses the prompt.
- **Goal:** Efficiency.
- **Warning:** Do not rely on factual assertions (dates, dollar amounts) without moving to Protocol B.

Protocol B: Substantive Review (The "Associate Review")

- **Action:** Treat the AI output as the work of a first-year associate. You assume it contains mistakes.
- **Checklist:**
  - Are client names spelled correctly?
  - Are dates consistent with the file?
  - Does the tone match firm standards?

Protocol C: Source Trace (The "Click-Through")

- **Context:** Used when the AI claims a fact is true based on a document (RAG).
- **Action:** You must click the citation link provided by the tool.
- **The Trap:** Does the highlighted text *actually* support the proposition? RAG tools often find the right case but hallucinate the *relevance* of a specific paragraph.
- **Requirement:** If the tool does not provide a clickable link to the source text, the output is inadmissible for work.

Protocol D: Forensic Audit (The "Zero Trust" Model)

- **Context:** Before any document is filed with a court or sent to an external party as final advice.
- **Action:** "Air-Gapped" Verification.
- **Process:** Take the citations generated by the AI. Open a *separate*, traditional legal database (Clean Browser Session). Manually search for the case/statute.
  1. Does it exist?
  2. Is it good law (shepardized)?
  3. Does it say what the AI said it says?

4. THE "SYCOPHANCY LOOP" AUDIT

Avoid Contributory Negligence by auditing your Prompts.

The Risk: AI models are trained to be helpful. If you ask a leading question, they will fabricate evidence to agree with you.

Prompt Self-Check:

- [ ] **Did I presuppose the answer?**
  - *Bad:* "Find cases where latency excuses breach of contract." (Forces AI to find a case, even if none exist).
  - *Good:* "Does server latency excuse breach of contract under NY law? Provide cases for and against."
- [ ] **Did I ask for a specific number of results?**
  - *Bad:* "Give me 5 cases." (If there are only 3 real cases, the AI may invent 2 to meet the quota).
  - *Good:* "List relevant case law."

5. VENDOR LIABILITY CHECKLIST (ISO 42001)

Questions to ask your Legal Tech Vendor before deployment.

1. **Grounding:** Does your RAG system have a "citation-only" mode that refuses to answer if no document is found?
2. **Indemnification:** Does your Terms of Service include an IP indemnity clause for generated content?
3. **Data Retention:** Is client data used to retrain your foundational models? (Must be NO).
4. **Insurance:** Do you carry Errors & Omissions (E&O) insurance that specifically covers AI "hallucinations" or failure of service?

Disclaimer: This document is for educational and governance purposes only. It does not constitute legal advice. Adherence to this matrix does not guarantee immunity from professional liability.