MPL LEGAL TECH ADVISORS
in collaboration with SARA PFROMMER, Esq.

The Legal AI Hallucination Guardrail Framework

SCIENCE BACKED, COURT TESTED AI PROMPT GUARDRAILS & VERIFICATION WORKFLOWS

Why We Created This Framework

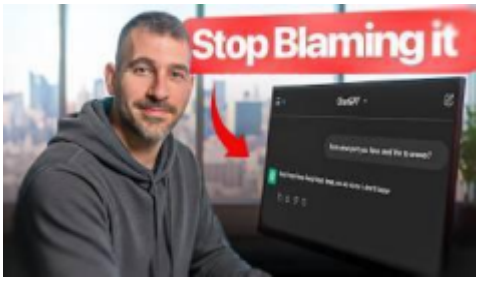
AI hallucinations are no longer a future concern, they are showing up in live court filings. Recent sanctions against attorneys at **Latham & Watkins**, **K&L Gates**, and **Butler Snow** [read here] prove that even top-tier firms can submit fake citations generated by ChatGPT.

- **Mata v. Avianca (S.D.N.Y. 2023):** Two New York lawyers were sanctioned **$5,000** after citing six fake cases generated by ChatGPT.
- **Johnson v. Dunn (N.D. Ala. 2025):** Two years later, sanctions are going further, **disqualifying** large-firm attorneys from an active case *and* referring them to the state bar for discipline after they submitted AI-invented case law.

What started as a monetary slap on the wrist is now career-level risk. The judiciary is signaling that "*I didn't know AI could do that*" **is no longer a defense**.

If you or your associates are using generative AI without controls, the next order with your name on it could be a **disqualification order, not just a fine**.

The two pillars of safe legal AI are Compliance and Evaluation. See the break-downs of both below:



[Why Will AI Keep Hallucinating]

[Legal AI Compliance Framework]

This Framework Combines:

- **Academic research:**
  - Why Language Models Hallucinate [read here]
  - Hallucination Is Inevitable: An Innate Limitation To Large Language Models [read here]
- **Real court cases:** Lessons from 370+ sanctioned lawyers [live case tracker]
- **Practical experience:** Co-developed with Utah appellate litigator Sara Pfrommer, Esq.

👉 *If you're unsure whether your firm's current AI use could expose you to compliance risks, book a 15min consultation. We'll review your workflows and flag any risks before they become sanctions.*

Book a 15min AI & Data Readiness Check

Part 1: The 11 Scientific Causes of AI Hallucinations

"Fake citations aren't random accidents, they're fundamental to how language models work."

Understanding *why* AI hallucinates is the first step to controlling it. Below are 11 scientific causes, translated into legal terms:

1.1. EPISTEMIC UNCERTAINTY

Legal translation: When AI encounters facts not in its training data

Example: Rare jurisdictional rules, recent case law updates

Risk: Fabricates plausible-sounding legal principles

1.2. TRAINING DATA IMPERFECTIONS

Legal translation: "Garbage in, garbage out" amplified

Example: Outdated statutes, incorrect legal summaries in training data

Risk: Perpetuates and amplifies existing legal misinformation

1.3. STATISTICAL PRETRAINING OBJECTIVES

Legal translation: AI predicts what sounds right, not what is right

Example: Generates citations that follow proper Bluebook format but cite non-existent cases

Risk: Perfect-looking but completely fabricated legal authorities

1.4. EVALUATION INCENTIVES FOR GUESSING

Legal translation: AI is trained to always answer, never admit uncertainty

Example: Providing confident-sounding but wrong legal advice rather than saying *"I don't know"*

Risk: Overconfident false statements in client communications

1.5. DISTRIBUTION SHIFT

Legal translation: When prompts differ from AI's training examples

Example: Novel legal scenarios, emerging technology law questions

Risk: Applies irrelevant legal precedents to new situations

1.6. MODEL ARCHITECTURAL DEFICIENCIES

Legal translation: Inherent limitations in AI's information processing

Example: Cannot properly track complex multi-party contract relationships

Risk: Misses crucial legal distinctions and nuances

1.7. ERRONEOUS DECODING AND INFERENCE

Legal translation: Word-by-word generation errors compound

Example: Starts with correct legal principle, but reasoning chain deteriorates

Risk: Internally consistent but externally invalid legal arguments

1.8. COMPUTATIONAL HARDNESS OF TASKS

Legal translation: Some legal problems are inherently difficult for any computer

Example: Complex regulatory compliance across multiple jurisdictions

Risk: Oversimplified solutions to complex legal problems

1.9. KNOWLEDGE OVERSHADOWING

Legal translation: Common legal knowledge drowns out specific exceptions

Example: Federal rules override specific state law exceptions

Risk: Misses jurisdiction-specific legal requirements

1.10. SNOWBALLING EFFECTS

Legal translation: One small error cascades into larger fabrications

Example: Incorrect case citation leads to fabricated holding and reasoning

Risk: Entire legal argument built on false foundation

1.11. SPURIOUS CORRELATIONS

Legal translation: AI connects unrelated legal concepts

Example: Links contract law with criminal procedure because they appear together in documents

Risk: Applies wrong legal framework to client situations

Every one of these causes is *predictable*, and **preventable**, once you know how to spot them. The next section walks through Judge Manasco's five-part taxonomy from *Johnson v. Dunn*; a practical map of how hallucinations break citations in real cases.

Part 2: The Five Citation Hallucination Types

"If fines and public embarrassment worked, we wouldn't still be seeing these cases."

| *Judge Manasco, Johnson v. Dunn (N.D. Ala. 2025)*

In July 2025, Judge Manasco laid out the clearest picture yet of how AI fails with legal citations. Her order identified **five distinct failure modes**, a playbook every lawyer should recognize before signing an AI-assisted filing.

TYPE 1: WRONG CASE, WRONG CITATION

- ***Example:*** "United States v. Baker, 539 F. App'x 937, 943 (11th Cir. 2013)"
- ***Reality:*** Real case exists but with different citation and unrelated content
- ***Prevention:*** Force-cite-before-answer prompts

TYPE 2: FABRICATED CITATION USING REAL CASE NAME

- ***Example:*** "Kelley v. City of Birmingham, 2021 WL 1118031"
- ***Reality:*** Case name exists but citation is completely fabricated
- ***Prevention:*** Closed universe prompts with verified databases

TYPE 3: COMPLETE FABRICATION

- ***Example:*** "Greer v. Warden, FCC Coleman I, 2020 WL 3060362"
- ***Reality:*** Neither case nor citation exists
- ***Prevention:*** Abstention prompts requiring uncertainty admission

TYPE 4: FABRICATED CITATION USING REAL CASE NAME

- ***Example:*** "Wilson v. Jackson, 2006 WL 8438651"
- ***Reality:*** Citation points to unrelated maritime case, not cited Rule 30 issue
- ***Prevention:*** Hard stop verification rules

TYPE 5: REAL CASE NAME, WRONG CITATION AND PROPOSITION

- ***Example:*** "Williams v. Asplundh Tree Expert Co., 2006 WL 3343787"
- ***Reality:*** Case style exists but not with that citation/holding combination

---

- ***Prevention:*** Multi-layered verification protocols

Each of these “failure modes” is controllable with the right safeguards, and in the next section, we show you the exact prompts and workflows that stop them before they reach a judge’s desk.

👉 *Before you buy another AI tool or risk citing the wrong case, book a consulting session. We’ll map your current risks and show you the simplest guardrails to put in place.*

Book a 15min AI & Data Readiness Check

Part 3: The 11 Science-Backed Prompt Templates

“In Johnson v. Dunn, three lawyers were disqualified for citing hallucinated cases”

Overarching Safety Features

“In Johnson v. Dunn, three lawyers were disqualified for citing hallucinated cases”

Here are some specific suggestions for how to handle the 11 identified elements of AI hallucination. In addition to these specific examples, these principles apply to every use of AI in legal context:

I. HUMAN JUDGEMENT IS ESSENTIAL

AI can scale your work and accelerate analysis, but it is not a substitute for legal judgment. Never file what you haven’t independently verified.

II. SEPARATE REASONING FROM SOURCING

Treat arguments and authority as distinct phases. Don’t allow the AI to blend them into a single output.

III. PUSH BACK CONSTANTLY

Treat the AI like a junior associate: press it to argue the other side, expose weaknesses, and admit uncertainty.

IV. KEEP PROJECTS IN SEPARATE THREADS

Maintain clean context. Don’t mix cases or issues in a single conversation, cross-contamination increases error risk.

V. LABEL BACKGROUND AS BACKGROUND

Distinguish non-legal context (policy, history, business practice) from binding legal authority.

CONTEXT MANAGEMENT GUARDRAIL

Problem addressed:
AI reasoning is only as good as the context it has. If you provide too little, the model will fill in the blanks with generic assumptions. If you provide too much, it may overweight irrelevant details. If you mix issues together, it may cross-contaminate. Each of these failures increases the risk of hallucination. You can minimize this concern by deliberately supplying structured context in the right order, and by requiring the AI to restate that context before applying it.

Suggested prompt language:
"Here is the context for this task:
- We represent [party].
- Our goal is [objective].
- The governing law/jurisdiction is [jurisdiction].
- The case is at [procedural posture].
- Relevant facts: [list].
- Constraints: [page limits, deadlines, strategic preferences].

Use only this context. Do not assume facts beyond what I have given you. Before analyzing, restate the context in your own words so I can confirm you are applying it correctly."

3.1. EPISTEMIC UNCERTAINTY GUARDRAIL

Problem addressed:
AI tends to invent confident-sounding but unreliable authorities when it encounters gaps in the law. You can minimize this concern by separating reasoning, research guidance, and analysis of real sources into distinct phases. Here are suggested prompts for doing that.

Phase 1 - Reasoning only:
"Give me your best reasoning about this issue, but do not include case names, citations, or statutes. Focus on the logic, doctrines, and potential arguments. If the law is unsettled, say so and explain why. It is the opposite of helpful if you give me an unreliable authority. If you need to assume any facts to frame the analysis, list those assumptions explicitly."

Phase 2 - Research guidance (not sources):
"Based on the reasoning above, suggest the kinds of sources where support might exist (e.g., state supreme court cases, recent appellate rulings, law review articles). Tell me which jurisdictions and date ranges are most likely relevant. Propose concrete search terms/filters I could use in Westlaw/Lexis. Do not create or quote specific citations - just help me focus my research."

Phase 3 - Analysis of identified sources
"Here are the real cases/statutes/articles I found. Please read and analyze them. Summarize the key holdings/rules, explain how they apply to the Phase 1 reasoning, and flag any conflicts or unsettled issues. If a source cuts against the reasoning, say so and explain the implications or how (if at all) it might be distinguished."

Disclaimer: *The Phase 1 and Phase 2 responses deliberately avoid providing specific case names or citations to prevent fabricated authorities. Always verify conclusions against real primary sources (statutes, case law, regulations) before relying on them.*

3.2. TRAINING CONTAMINATION GUARDRAIL

Problem addressed:
While Epistemic Uncertainty (Prompt #1) deals with the risk of AI inventing new, unreliable authorities to fill gaps. Training Contamination is about AI repeating bad, outdated, or incorrect information it absorbed during training and presenting it as if it were binding law. You can minimize this concern by making the AI separate background knowledge from verifiable sources, and by explicitly instructing it not to rely on its training data as authority.

Suggested prompt language:
"Do not rely on your training data as proof of a legal rule. If you mention general legal principles from your training, label them clearly as background, not authority. Only treat the reasoning I supply or real sources I provide as authoritative."

Disclaimer: *This assistant's training data is pre-2025 and not authoritative. Treat any principles mentioned as background only and confirm against current primary sources before relying on them.*

3.3. STATISTICAL PREDICTION GUARDRAIL

Problem addressed:
Even when not inventing or repeating bad data, AI may default to what sounds right - the statistically likely answer - instead of the jurisdiction-specific or legally accurate one. You can minimize this concern by requiring the AI to disclose its level of confidence and explain the basis for that confidence, rather than presenting every answer as equally certain.

Suggested prompt language:
"For each legal rule or proposition you provide, assign a confidence level (high / medium / low) and explain why you chose it. If your confidence is less than high, tell me what factors create uncertainty - for example, differences among jurisdictions, unsettled precedent, or lack of clarity in the law. Do not smooth over these distinctions. Present the most jurisdiction-specific rule first, then broader generalizations."

Disclaimer: *Confidence levels are heuristic and not legal authority. Confirm jurisdiction-specific rules and primary sources before relying on any statement.*

3.4. UNCERTAINTY ADMISSION GUARDRAIL

Problem addressed:
While Epistemic Uncertainty (Prompt #1) addresses the risk of AI inventing authority to fill a gap, Uncertainty Admission is about candor: the AI's tendency to pretend to know even when the law is unsettled, contested, or outside its depth. Instead of openly acknowledging uncertainty, it often presents a confident answer that hides the lack of clarity. You can minimize this concern by requiring the AI to admit when it does not know and to explain the reason for the uncertainty.

Suggested prompt language:
"Do not pretend to know. If you are uncertain, say so directly: 'This is

unsettled law' or 'I do not know the answer with confidence.' Then explain why - for example, courts are split, the statute is new, or precedent is thin. If there are competing interpretations, list them without choosing one as the final answer."

Disclaimer: *This output may highlight unsettled law. Confirm all flagged uncertainties with primary sources before relying on them.*

3.5. DISTRIBUTION SHIFT GUARDRAIL

Problem addressed:
AI is trained on past data and often applies old patterns to new situations. When it encounters a novel statute, emerging area of law, or unusual fact pattern, it may mislead you by treating unsettled issues as if they were well-established. You can minimize this concern by requiring the AI to classify first whether an issue is settled law or emerging/novel law, and to explain that classification before giving its reasoning. (See also Prompt #12, Active Pushback, which reinforces the need to press the AI on whether an issue is genuinely settled.)

Suggested prompt language:
"Before answering, classify this issue: is it well-established law, or emerging/novel law? Explain why you put it in that category. If it is emerging or novel, identify what makes it uncertain (e.g., new statute, sparse precedent, split of authority) and give me the range of possible interpretations rather than a single answer."

Disclaimer: *This classification flags whether an issue may be unsettled. Confirm with primary sources before relying on the analysis.*

3.6. ARCHITECTURAL LIMITATION GUARDRAIL

Problem addressed:
Some legal questions are too complex to be answered accurately in one sweep. An AI may compress or oversimplify multi-factor analyses, overlapping statutes, or fact-heavy disputes, presenting a tidy answer where none exists. You can minimize this concern by requiring the AI to first identify whether the problem is sufficiently complex to be broken into smaller pieces, and if so, to break the problem into smaller parts

that can be addressed one at a time.

Suggested prompt language:
"Before answering, identify whether this problem is sufficiently complex that it should be broken into smaller parts. If so, explain what makes it complex (e.g., multiple parties, overlapping statutes, fact-intensive issues, or procedural layers). Then break the problem into smaller parts that can be addressed one at a time, rather than all at once, and flag where human legal judgment is required."

Disclaimer: *This output may oversimplify complex issues. Treat each part as a starting point and confirm with primary sources before relying on the analysis.*

3.7. ERROR PROPAGATION (CASCADE and CHAIN) GUARDRAIL

Problem addressed:
Once the AI makes a mistake, that error can spread in two ways: (a) a single bad assumption at the start can cascade through everything that follows, or (b) a weak link in the middle of a reasoning chain can undermine later steps. Either way, the result is a long, coherent, but ultimately flawed analysis. You can minimize this concern by requiring the AI to pause and restate its assumptions before building on them, and by validating each link in a reasoning chain independently. The best guardrail is still human judgment: a lawyer must review each step to ensure the foundation is solid before moving to the next one.

Suggested prompt language:
"Before building out a full argument, list the key assumptions or premises you are relying on. For each, explain your level of confidence. Then pause. Do not continue unless I confirm that these premises are sound. Remember: if any assumption is incorrect, everything that follows may be flawed. Treat each reasoning step as a separate link in the chain, and flag it for human review before moving on."

Disclaimer: *This step-by-step output is designed to catch errors before they spread. Verify each paused element with primary sources before continuing.*

3.8. COMPUTATIONAL HARDNESS GUARDRAIL

Problem addressed:

Some legal problems involve so many moving parts (family trees, asset distributions, overlapping statutes) that the analysis can collapse under its own weight. AI is well-suited to handle the mechanical untangling — organizing relationships, mapping scenarios, running calculations — but it is not equipped to decide the legal consequences without human review. You can minimize this concern by instructing the AI to separate structural mapping from legal interpretation.

Suggested prompt language:
"If this problem is computationally complex (many permutations, branching family trees, overlapping statutes, or layered timelines), say so. First, untangle the structure step by step - e.g., list the family relationships, the sequence of events, or the distribution of shares - without interpreting the legal effect. Then pause. Only after the structural map is clear should we move to legal interpretation, which requires human judgment. Flag explicitly any parts where law or policy, not computation, must decide the outcome."

Disclaimer: *This output maps complex structures but does not decide legal outcomes. Verify each step and document assumptions before relying on the interpretation.*

3.9. KNOWLEDGE PRIORITIZATION GUARDRAIL

Problem addressed:
Because legal sources make up only a small slice of AI's training data, the model may default to broad generalizations or colloquial background instead of the narrow, jurisdiction-specific rules that actually govern. The key safeguard here is to make the order of operations explicit: controlling law must come first, broader principles second, and any non-legal background only after that and clearly labeled as such.

Suggested prompt language:
"Start with the most specific, jurisdiction-relevant legal rule or definition. Present that first. Only after you have given the controlling legal standard may you add broader principles or colloquial background - and if you do, label that material clearly as background, not authority. Do not blur colloquial usage with legal rules."

Disclaimer: *This output prioritizes controlling law first, then general principles. Verify cited authorities and confirm jurisdiction before relying on the conclusion.*

3.10. CORRELATION vs. CAUSATION GUARDRAIL

Problem addressed:
AI may confuse concepts that often appear together in legal texts with concepts that are legally linked. For example, fraud claims often travel alongside breach of contract, but proving breach does not establish fraud. You can minimize this concern by requiring the AI to separate correlation from causation and to say explicitly whether two concepts are legally connected by rule, statute, or doctrine, or whether they merely co-occur in practice.

Suggested prompt language:
"Do not assume that two concepts are legally connected just because they often appear together. Tell me whether they are linked by statute, doctrine, or precedent, or whether they merely correlate in practice. Make the distinction explicit in your answer."

Disclaimer: *This output distinguishes legal connection from mere co-occurrence. Confirm cited authorities to ensure the link is valid in your jurisdiction.*

3.11. ACTIVE PUSHBACK GUARDRAIL

Problem addressed:
Even with guardrails in place, AI tends to sound authoritative and may overstate one side of an argument. Like a human associate, it may hesitate to volunteer the weaknesses in its own reasoning. You can minimize this concern by actively pushing back: requiring the AI to argue the other side, surface weaknesses, and identify assumptions that might not hold.

Suggested prompt language:
"You've told me our best argument is [X]. Now argue the other side as strongly as you can.""Identify the weaknesses in the position you just gave me. If you were opposing counsel, how would you attack it?""List the assumptions you made in your reasoning. What happens if those assumptions turn out to be wrong?"

Disclaimer: *This output surfaces counterarguments and weaknesses for review. Confirm each point with primary sources before relying on the analysis.*

Illustrative Example

#5: DISTRIBUTION SHIFT - CRYPTO LENDING BUSINESS

This example shows how answers differ with and without the guardrail in place.

Scenario: A client wants to form a business to lend crypto to borrowers.

Without guardrail:
"If your client wants to form a business to lend crypto to borrowers, the main legal concerns include securities laws (crypto assets may be securities under the Howey test), lending and usury regulations, money transmission and AML/KYC requirements, tax treatment, and bankruptcy risks. You should ensure compliance with SEC and CFTC regulations, obtain any necessary state lending or money transmitter licenses, and establish robust consumer protection and risk disclosure practices."

NOTE:

- Reads smoothly and confidently, as if the law were settled.
- Blends background knowledge with legal authority without distinction.
- Omits the fact that much of the law is still unsettled.

With guardrail:
"This is emerging**/novel law**.

Why:
Crypto lending implicates evolving federal guidance (SEC, CFTC, OCC), unsettled court precedent, and new state statutes. There is no fully settled body of law comparable to traditional lending.

Sources of uncertainty:
Federal–state regulatory split; contested definition of tokens as securities; recent enforcement actions without appellate precedent; unsettled bankruptcy treatment of customer assets; incomplete IRS guidance.

Range of possible issues:
Securities laws, state lending/usury rules, money transmission laws, commodity rules, bankruptcy classification of assets, and tax reporting concerns.

Bottom line:
This is unsettled territory. Here's why it's unsettled, and here are the risks regulators and courts are likely to focus on."

NOTE:

- Flags novelty and uncertainty explicitly.
- Explains why the issue is unsettled.
- Maps the range of competing risks instead of presenting a single "answer."

Contrast:
Without the guardrail, the output looks neat and authoritative but conceals how unsettled the law really is. With the guardrail, the novelty is surfaced, uncertainty is explained, and the risks are mapped transparently.

Disclaimer: *This output distinguishes legal connection from mere co-occurrence. Confirm cited authorities to ensure the link is valid in your jurisdiction.*

👉 *If you want your firm to have this level of clarity before filing anything with a judge, book a consulting session. We'll walk you through the exact process tailored to your practice.*

Book a 15min AI & Data Readiness Check

Sara Pfrommer's Hallucination Mining Blueprint

Sara Pfrommer, an appellate litigator in Utah, has developed a structured way to turn AI hallucinations into useful prompts for deeper legal research.

Rather than treating hallucinations as purely negative, she treats them as starting points for exploring novel arguments while still verifying every fact and citation.

THE 2 STEP PROCESS

STEP 1: ARGUMENT-ONLY EXTRACTION

This step isolates the arguments' logical core before sourcing authority. The goal is to surface the strongest possible theory without risking fabricated citations.

Sara's prompt framework:

```
You are a legal analysis assistant. Your task is to develop the
strongest possible legal argument for [LEGAL POSITION].

CRITICAL RESTRICTIONS:
- Do NOT cite cases, statutes, or regulations.
- Do NOT provide specific legal authorities.
- Do NOT mention dates, jurisdictions, or case names.
- If you include a citation, it will be disregarded.

FOCUS ON:
- Core legal principles and reasoning
- Logical argument structure
- The theoretical framework that supports the position
- Key counterarguments and how to address them
- The doctrinal steps that would succeed if properly supported by
authority

EXAMPLE OUTPUT:
"""
The strongest argument rests on the principle that [LEGAL THEORY]. This
framework suggests that [LOGICAL REASONING]. The key doctrinal move is
[STRATEGIC APPROACH]. Potential counterarguments include [OPPOSING
VIEWS], which can be addressed by [RESPONSE STRATEGY].
"""
```

Why it works:

- ➔ **Scientific grounding:** Sidesteps epistemic uncertainty and training data imperfections
- ➔ **Lawyer's advantage:** Isolates theory of the case and crucial elements
- ➔ **Risk elimination:** No fake citations can sneak into your draft

STEP 2: HUMAN SOURCING

Once Sara has the argument, she puts on her lawyer hat and runs targeted searches to find real authority.

Verification questions:

- ➔ **Holding vs. dicta:** Is the case actually deciding the issue, or just discussing it?
- ➔ **Procedural posture:** Trial or appeal? What's the standard of review?
- ➔ **Treatment & jurisdiction:** Is the case still good law? Binding or persuasive?
- ➔ **Fact alignment:** Does the authority fit your actual case scenario?

Research strategy:

1. Take the AI-generated argument theory
2. Run targeted Westlaw/Lexis searches using the theoretical framework
3. Look for cases that actually support the principle
4. Verify jurisdiction, treatment, and factual alignment
5. If no solid source found quickly, reframe or drop the argument

If AI suggests a case name:

- ➔ Treat it as a research lead, never as authority
- ➔ Never trust it until independently verified
- ➔ Use the suggested case name as a starting point for broader research

Advanced Prompt Engineering for Legal Professionals

"Courts now expect a defensible verification process, not just promises you'll be careful."

THE SCOPE METHOD FOR LEGAL PROMPTING

Prepare a one-page assessment summary for each AI tool:

S - SPECIFY THE LEGAL CONTEXT

- ➔ Practice area, jurisdiction, court level
- ➔ Procedural posture and timeline
- ➔ Relevant legal standards

C - CONSTRAIN THE AI'S ROLE

- ➔ Research assistant, not legal advisor
- ➔ Cite limitations and verification requirements
- ➔ Define output format and structure

O - OUTLINE EXPECTED DELIVERABLES

- ➔ Specific work product needed
- ➔ Level of detail required
- ➔ Format for citations and authorities

P - PROVIDE COMPREHENSIVE BACKGROUND

- ➔ Relevant facts (redacted appropriately)
- ➔ Legal issues already identified
- ➔ Strategy considerations

E - ESTABLISH VERIFICATION PROTOCOLS

- ➔ Relevant facts (redacted appropriately)
- ➔ How to flag uncertainty
- ➔ Process for source checking

4 Layer Prompt Builder

FOUNDATION LAYER

```
You are a [SPECIFIC ROLE] assistant specializing in [PRACTICE AREA] for
[JURISDICTION].

Your role is limited to [SPECIFIC TASKS].

You must [VERIFICATION REQUIREMENTS].
```

CONTEXT LAYER

```
Current matter involves [CASE CONTEXT].

Key legal issues: [ISSUE LIST]

Procedural status: [STATUS]

Strategic considerations: [STRATEGY]
```

TASK LAYER

```
Specific task: [DETAILED REQUEST]

Required format: [FORMAT SPECIFICATIONS]

Output limitations: [WHAT NOT TO INCLUDE]

Quality standards: [VERIFICATION CRITERIA]
```

SAFETY LAYER

```
If uncertain about [SPECIFIC AREAS], respond with: [UNCERTAINTY FORMAT]

Required verification steps: [VERIFICATION LIST]

Escalation triggers: [WHEN TO STOP AND ASK FOR HELP]
```

Quick Reference: Summary

1. **Know the stakes:** sanctions now span fines (Mata v. Avianca, 2023) to disqualification and bar referral (Johnson v. Dunn, 2025).
2. **Use the 5 mode check:** wrong case/wrong cite; real name/fake cite; full fabrication; misdirected cite; wrong cite + proposition.
3. **Separate tasks:** have AI draft arguments only; block and strip AI-inserted citations.
4. **Verify sources:** confirm holdings in trusted databases, then insert only verified authorities with links.
5. **Map cause - control:** apply the 11 cause-matched prompts (uncertainty, training flaws, decoding errors, etc.).
6. **Require abstention:** use uncertainty/"source needed" formats and confidence labels before conclusions.
7. **Break error chains:** pause on unverified elements; resume only after primary authority is confirmed.
8. **Prioritize specific law:** local/jurisdiction rules before general principles; classify novel/edge issues first.
9. **Use Sara Pfrommer's blueprint:** extract argument without cites > human sourcing with holding/pose, posture, treatment, fit.
10. **Standardize prompting:** SCOPE method + 4-layer prompt builder (Foundation, Context, Task, Safety) for each matter.

Conclusion: From Evaluation to Excellence

AI can be used safely in legal work if you separate drafting from citation, verify sources, and encode the workflow steps that don't change by case.

This guide provides:

- **Evidence base:** 11 documented causes of hallucinations with matching controls.
- **Tools:** prompt patterns, checklists, and validation steps used in practice.

- **Method:** Sara Pfrommer's hallucination mining approach to find and fix weak points.

Sara's golden rule: *"If we haven't read it, we can't vouch for it. And if we can't vouch for it, it doesn't belong in our brief."*

If you don't have the time to build this yourself, this is exactly what we do at MPL Legal Tech Advisors. Book your consulting session, and we'll create a safe, compliant AI roadmap for your firm.

Book a 15min AI & Data Readiness Check
