SMBTeam

LAW FIRM AI SAFETY

The AI Hallucination Guard Playbook

How to use Second Opinion Mode to catch fake citations before they reach a judge

There's a public database of AI mistakes that has your profession's name on it. As of October 5, 2026, it logs more than 2,100 court cases where AI fabricated legal citations. Most invented cases that don't exist. Some quoted real cases for things they never said. Judges are sanctioning lawyers for it.

This playbook shows you how to use AI at full speed without putting your license on the line.

2,100+

CASES LOGGED, AS OF OCT 5, 2026

AI hallucination cases logged in a public court database.

100%

PLANTED FAKES CAUGHT

Second Opinion Mode found 2 of 2 planted fake citations in a live test.

Why AI hallucinates, and why it matters 2

What Second Opinion Mode does 3

How to use it, step by step 4

What the output looks like 5

It does not replace your review 6

Getting started with AI Workforce Pro 7

THE PROBLEM

01 Why AI hallucinates, and why it matters for lawyers

The problem isn't that AI is unreliable. It's that it's confidently wrong.

AI language models generate text by predicting what comes next, not by looking things up in a verified database. When a model is asked about a case it hasn't seen, it may produce a plausible-sounding citation rather than admitting it doesn't know. The result looks exactly like a real citation, with the correct format, a credible court, and a realistic year. But the case doesn't exist.

There are three ways this shows up in legal documents:

HALLUCINATION TYPE	WHAT IT LOOKS LIKE	RISK TO YOUR LICENSE
Invented case citations	AI cites a case name, court, and year that don't exist	Sanctions, contempt, potential bar discipline
Fabricated quotes	AI attributes a quote or holding to a real case that never said it	Opposing counsel catches it in court and your credibility takes an immediate hit
Misattributed authority	Real case cited for a proposition it never established	Silently undermines your argument and may not be caught until it matters most

> **WHY THIS IS DIFFERENT FROM A RESEARCH ERROR**
>
> A human researcher who cites the wrong case usually cites a real case, one that exists but supports a different point. AI hallucinations create citations that don't exist at all. Opposing counsel or a clerk trying to pull the case finds nothing.
>
> **That's a different category of error, and courts are treating it differently.**

THE SAFEGUARD

02 What Second Opinion Mode does

Two models checking independently makes a shared hallucination nearly impossible

Second Opinion Mode is a feature inside AI Workforce Pro that sends the same document and prompt to two AI models simultaneously and independently. Neither model sees the other's output while it's working. At the end, you get a consensus summary that shows what both found and flags anything either model questioned.

> **WHY TWO MODELS BEAT ONE**
>
> It has no way to catch its own blind spots. Two models trained differently on different data would have to independently invent the exact same fabricated case, with the same name, court, and year, for the hallucination to slip through.
>
> **If a single model hallucinates a case citation, that same model is likely to confirm it when asked to verify its own work.**
>
> **The probability of that is close to zero.**

MEASURE	SINGLE MODEL ONLY	SECOND OPINION MODE
Hallucination risk	Model's own blind spots unchecked	Near zero, since two models inventing the same fake case is nearly impossible
Citation verification	Model checks its own output	Two independent models cross-check each other
Output format	Raw response that needs manual review	Consensus summary plus individual full responses
Time required	Fast, one pass	Slightly longer, two parallel passes, then consensus
Confidence before filing	Moderate, still needs a manual check	High, plus your own final review

THE WORKFLOW

03 How to use it, step by step

Six steps from document upload to verified citations

AI Workforce Pro puts ChatGPT, Claude, and Gemini in one place. Here's exactly how to run a citation verification using Second Opinion Mode.

1 Open AI Workforce Pro

Log in, where ChatGPT, Claude, and Gemini are all available in a single interface. Select the primary model you want to use, either the one that generated the original document or the one you want for the first pass of research.

2 Upload your document

Upload the brief, motion, memo, or any document containing citations you want to verify, then use the verification prompt below.

3 Click Get a Second Opinion

Click the second opinion button (the squares icon). This sends the same document and prompt to two additional models simultaneously and independently.

4 Select two verification models

Choose two models different from the one that created the content. If ChatGPT wrote the document, select Gemini and Claude Sonnet. Neither sees the other's output during review.

5 Review the combined findings

When both models finish, you receive a consensus summary of what they found together. Click Full response to see each model's individual findings. The consensus highlights any citation flagged by either model.

6 Remove or verify flagged citations

Any citation flagged as fabricated must be verified manually or removed before filing. Do not file until every flag is resolved.

THE EXACT VERIFICATION PROMPT TO USE

I'm uploading a document. I want you to verify all the citations are accurate and real.

For each citation, confirm:
1. The case exists and can be found in legal databases
2. The court and year are correct
3. The case actually supports the proposition cited

Flag any citation you cannot confirm as UNVERIFIED. List verified citations separately from unverified ones.

THE PROOF

04 What the output looks like: a real example

Five legitimate, two fabricated, caught before filing

In a live test, a document was prepared with five real Supreme Court citations and two deliberately fabricated federal appellate citations. Second Opinion Mode using Gemini and Claude Sonnet returned this combined finding:

COMBINED FINDINGS FROM THE SECOND OPINION MODE TEST

"Both responses agree that five citations cited the Supreme Court cases. They also agree that two federal appellate cases are fabricated citations that must be removed and replaced with real legal authority before filing."

Result: 5 legitimate confirmed. 2 fabricated detected. 0 missed.

What to do with the output

- Read the consensus summary first. It shows what both models agreed on
- Click Full response for each model's individual findings if you want detail
- Any citation flagged UNVERIFIED by either model must be manually checked or removed
- Any citation flagged by both models should be treated as fabricated until proven otherwise
- Do not file until every flag in the consensus has been resolved

THE LIMITS

05 It does not replace your review

It raises your confidence. It doesn't remove your responsibility.

Second Opinion Mode is hallucination insurance, not hallucination immunity. It dramatically reduces the probability that a fabricated citation makes it into your filing. It doesn't eliminate the need for your own review. Here's what it covers and what it doesn't.

What Second Opinion Mode covers well

- Invented case citations, meaning cases that don't exist at all
- Fabricated quotes attributed to real cases
- Obvious misattributions, such as a real case cited for a holding it never made
- Any citation that both models can't independently verify

What still requires your own review

- Subtle misreadings of a real case, where the case exists but the nuance is wrong
- Jurisdiction-specific precedent where model training data may be thin
- Very recent cases published after the models' knowledge cutoffs
- Your own independent judgment on how a case applies to your specific facts

> THE RIGHT WORKFLOW
>
> Verify every citation before filing, then use the consensus to decide what to pull and read first.
>
> **Citations flagged by both models are urgent. Citations flagged by one are worth a closer look. Citations cleared by both still deserve your review, but you can approach them with significantly higher confidence.**

NEXT STEPS

06 Getting started with AI Workforce Pro

One tool, all three major models, built for law firms

AI Workforce Pro puts ChatGPT, Claude, and Gemini in one unified interface designed for law firms. Second Opinion Mode is built in, with no additional setup required.

Three things to do this week

- Run one document you recently drafted through Second Opinion Mode before the next filing deadline. Start with any brief that cites more than three cases
- Bookmark the AI hallucination court case database and check whether any cases you've recently cited appear on the list
- Share this playbook with anyone at your firm who uses AI for legal research. The risk isn't limited to partners

SMBTeam

Scale your law firm across leads, intake, operations, and finance

Learn more