SMBTeam

LAW FIRM AI SAFETY

The Law Firm Citation Verification Playbook

How to catch AI-hallucinated citations before they end up in a court filing

A 60-second cross-model check that gives you confirmation instead of assumption.

2,100+
DOCUMENTED CASES

AI hallucinations have now shown up in more than 2,100 real court filings. A hallucinated citation has a case name, a court, a year, and a page number. The only problem is that **it does not exist.**

Judges and opposing counsel have started checking. Lawyers have faced sanctions, bar complaints, and public embarrassment when they didn't.

Why AI hallucinates citations 2

The fix: cross-model verification 3

How to run a citation check 4

What the results look like 5

When to run a citation check 6

AI as a team, not a single source 6

Quick reference 7

THE PROBLEM

01 Why AI hallucinates citations, and why you can't always spot it

The problem isn't that the output looks wrong. It's that it looks exactly right.

When you ask Claude, ChatGPT, or Gemini to help research or draft a legal document, the AI doesn't look up cases in a legal database. It generates text based on patterns in the data it was trained on. Most of the time, that produces accurate, useful work. But with specific citations, like case names, docket numbers, court designations, and page numbers, the model sometimes fills gaps with plausible-sounding fabrications.

The dangerous part is that hallucinated citations don't look suspicious. They look authoritative. They follow the correct citation format. They reference courts that exist. The case name sounds like something a real court would produce. You'd only know it was fabricated by checking Westlaw, Lexis, or a court database, which takes time and is easy to skip under deadline.

> SELF-CHECK PROBLEM
>
> **Why the model that wrote the document can't check itself.**
>
> **If Claude wrote your brief and you ask Claude to verify its own citations, you're asking the same system that generated the hallucination to detect it.**

That model doesn't have access to a real legal database. It can't look up whether a case actually exists. What it can do is confirm the citation sounds reasonable, which it already believed when it wrote it. Self-verification doesn't work.

THE FIX

02 Cross-model citation verification

Two independent AI models review the same document, and neither knows what the other found.

The solution isn't to stop using AI. It's to verify AI output using multiple independent models. The odds that Claude hallucinates a specific fake citation **and** ChatGPT or Gemini independently hallucinate the exact same fake are near zero. The models are trained differently, on different data, with different architectures.

When they agree on a citation, that agreement means something. When they disagree, you know exactly where to look.

TRUSTING ONE AI MODEL ALONE	USING SECOND OPINION MODE
One model checks its own work and misses the errors it made.	Two independent models review the same document without knowing what the first produced.
No consensus. You have one answer and no way to verify it.	Agreement signals confidence. Disagreement flags exactly where to look.
A hallucinated case sounds exactly like a real one in the output.	If one model fabricates a citation, the other is very unlikely to fabricate the same one.
You catch the problem during a hearing or opposing counsel's review.	You catch the problem in 60 seconds, before the document leaves your desk.
Over 2,100 AI hallucination cases are documented in real court filings.	Cross-model verification makes the same-error overlap near impossible.

THE PROCESS

03 How to run a citation check in AI Workforce Pro

Three steps. About 60 seconds.

AI Workforce Pro's Second Opinion Mode lets you send any document to multiple AI models at the same time and see both their individual responses and a combined summary. No switching tabs. No copying and pasting prompts.

#	STEP	WHAT HAPPENS
1	**Upload your document**	Drag your brief, motion, or legal memo into AI Workforce Pro. It reads the full document and identifies every citation inside it: cases, statutes, regulations, and secondary sources.
2	**Ask for a second opinion**	Tell it you want every citation verified as accurate and real. AI Workforce Pro sends the document and your request to ChatGPT and Gemini at the same time, two models that didn't write the original.
3	**Review the combined findings**	Both models review every citation independently. You get each model's individual response plus a combined summary showing where they agree and where they differ, so you know which citations are confirmed and which need manual review.

THE EXACT PROMPT TO USE

"I'm uploading a document. I want you to verify all the citations are accurate and real. Review every case, statute, and legal authority cited in this document. Flag any citation that you cannot verify as legitimate, and explain why you are flagging it."

THE PROOF

04 What the results actually look like

A live demo with 5 real citations and 2 planted fakes

To show how Second Opinion Mode works, we prepared a document with five authentic Supreme Court citations and two completely fabricated federal circuit court citations. We uploaded it to AI Workforce Pro, submitted the prompt above, and ChatGPT and Gemini reviewed it independently.

What the combined findings showed

- Both ChatGPT and Gemini independently confirmed the five authentic Supreme Court citations as real
- Both models flagged the same two fabricated federal circuit citations as unverifiable
- The combined summary clearly separated confirmed citations from flagged ones
- Neither model was told what to look for. They found the fakes without any hints

The combined view shows exactly where the two models agree (high confidence, the citation is likely real) and where they disagree or flag issues (review required). You don't have to read two full responses. The summary tells you what you need to know.

> **WHEN A CITATION IS FLAGGED**
>
> It means at least one model couldn't verify it. Check it manually in Westlaw, Lexis, Google Scholar, or the court's own docket system.
>
> A flag doesn't automatically mean the citation is hallucinated.
>
> **If the case exists, the flag is cleared. If it doesn't, you caught a hallucination before it reached a court filing.**

THE WORKFLOW

05 When to run a citation check

Build it into your workflow, not as an afterthought.

Citation verification is most valuable right before a document leaves your desk. Make it the last step before you finalize any AI-assisted filing. You can also use it at the research stage to validate cases the AI surfaces before you build arguments around them.

Documents to run through Second Opinion Mode before filing

- Motions and briefs with AI-assisted research or drafting
- Demand letters citing case law
- Legal memoranda referencing statutory authority or precedent
- Any document where opposing counsel or a judge will verify citations
- Research summaries produced by AI that you plan to rely on

Also useful at the research stage

- When AI surfaces a case you don't immediately recognize, run it through Second Opinion Mode before building your argument around it
- When you're researching a niche area of law where you know fewer of the key cases
- Any time you're under deadline pressure and want a fast confidence check before manual verification

THE BIGGER PICTURE

06 AI as a team, not a single source

No single AI model should be your only check on AI-generated work. You wouldn't rely on one expert witness without corroboration, so don't rely on one AI model to verify work another AI model produced.

> **THE THREE-MODEL PRINCIPLE**
>
> When all three agree, you have strong signal. When they disagree, you have a specific flag.
>
> Claude, ChatGPT, and Gemini are different systems with different training data and different blind spots.
>
> **The goal isn't to avoid AI. It's to use it with the same professional diligence you bring to everything else.**

CHEAT SHEET

07 Citation check quick reference

Everything you need in one place.

The problem

- AI models generate plausible-sounding citations that don't exist
- Over 2,100 documented cases of AI hallucinations in real court filings
- The model that wrote the document can't reliably verify its own citations

The solution

- Use AI Workforce Pro's Second Opinion Mode to send your document to ChatGPT and Gemini at the same time
- Both models review every citation independently, and neither sees what the other found until the summary
- The combined findings view shows confirmed citations vs. flagged ones in plain language

The prompt

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

When to use it

- Before filing any document with AI-assisted research or drafting
- When AI surfaces a case you don't immediately recognize
- Any time you want confidence without spending an hour in Westlaw first

SMBTeam

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

Learn more