

Citation checking report, cutoff Oct 26, 2026

As of Oct 26, 2026, aiforlawyers.app counts 290 lawyer voices on citation checking; 246 voices name one of 12 tools, 4 of which have 10+ voices and are rated. Highest floor of the 90% band: Claude (net +16, band -12 to +41, 31 voices, Low confidence).

Online version, data and updates: <https://aiforlawyers.app/report/citation-checking>

CITE THIS

aiforlawyers.app (AI Vortex), Citation checking report, cutoff Oct 26, 2026,
<https://aiforlawyers.app/report/citation-checking>, licensed CC BY 4.0.

For one figure: “As of Oct 26, 2026, aiforlawyers.app counts 290 lawyer voices on citation checking; 246 voices name one of 12 tools, 4 of which have 10+ voices and are rated. Highest floor of the 90% band: Claude (net +16, band -12 to +41, 31 voices, Low confidence).”

Every tool, with its voices

12 tools named in 290 voices on citation checking. Rated tools first, ordered by the floor of the 90% band; then the tools with few voices, with their count and no band. Net = like minus dislike per 100 voices.

#	Tool	Status	Voices	Net	90% band	Confidence	Top complaint (voices)
1	Claude	Rated	31	+16	-12 to +41	Low	Hallucination (10)
2	Lexis+ AI	Rated	17	-47	-71 to -9	Low	Hallucination (6)
3	ChatGPT	Rated	157	-69	-76 to -59	Medium	Hallucination (111)
4	Westlaw AI	Rated	11	-45	-77 to +3	Low	Hallucination (5)
-	Gemini	Few voices	9	-	no band: 9 of 10 voices needed	not rated	Hallucination (7)
-	CoCounsel	Few voices	7	-	no band: 7 of 10 voices needed	not rated	Hallucination (2)
-	Harvey	Few voices	6	-	no band: 6 of 10 voices needed	not rated	Hallucination (4)
-	Perplexity	Few voices	3	-	no band: 3 of 10 voices needed	not rated	Hallucination (1)
-	Copilot	Few voices	2	-	no band: 2 of 10 voices needed	not rated	Hallucination (1)
-	Legora	Few voices	1	-	no band: 1 of 10 voices needed	not rated	none on record
-	Spellbook	Few voices	1	-	no band: 1 of 10 voices needed	not rated	none on record

#	Tool	Status	Voices	Net	90% band	Confidence	Top complaint (voices)
-	vLex Vincent	Few voices	1	-	no band: 1 of 10 voices needed	not rated	Hallucination (1)

What lawyers complain about, in their words

The most named complaints per tool, with a verbatim quote each. Counts are voices, not opinions of the firm or of us.

Claude Rated · 31 voices

15 voices of 31 criticize it or are mixed about it with a named complaint: hallucination (10), accuracy (2), trust, liability (2), price (1).

“I've built in every conceivable safeguard that I can, and Claude's model just **consistently* underwhelm.*”

r/legaltech · Jun 8, 2026 · Hallucination

“Lawyers representing #AI firm Anthropic in a lawsuit inadvertently put citation errors in their brief after using Anthropic's own tool, Claude, to format the material.”

Bluesky · Jun 2, 2025 · Accuracy

Lexis+ AI Rated · 17 voices

12 voices of 17 criticize it or are mixed about it with a named complaint: hallucination (6), trust, liability (4), price (1), integration (1).

“When I tried to Sheperdize the cite, it became apparent this little nugget was cut from whole cloth.”

r/Lawyertalk · Jul 5, 2026 · Hallucination

“What I find are real cases but misquoted and irrelevant. Protege is very expensive and a huge rip off.”

r/Lawyertalk · Jul 6, 2026 · Price

ChatGPT Rated · 157 voices

132 voices of 157 criticize it or are mixed about it with a named complaint: hallucination (111), accuracy (10), trust, liability (10), privacy (1).

“Duo from London law firm self-report to the SRA after filing fake quotes generated by ChatGPT and 'infected' with AI”

Bluesky · Oct 9, 2026 · Hallucination

“ChatGPT isn't capable of that kind of research yet.”

r/Lawyertalk · Dec 31, 2025 · Accuracy

Westlaw AI Rated · 11 voices

8 voices of 11 criticize it or are mixed about it with a named complaint: hallucination (5), accuracy (3).

“In the one case the client incorporated the changes in redline and mentioned that they were from Westlaw AI so I checked the citations and found a couple were inaccurate or misleading and (gently) told her that.”

r/Lawyertalk · Jul 21, 2026 · Hallucination

“Running a brief through Westlaw cite check is insufficient IMO. It is just another AI platform.”

r/Lawyertalk · Aug 19, 2026 · Accuracy

Gemini Few voices · 9 of 10

7 voices of 9 criticize it or are mixed about it with a named complaint: hallucination (7).

“Richard Ervin, the Attorney General of Florida, is supposed to have died in 1962, collapsing while arguing the case of Gideon v. Wainwright in front of SCOTUS. Fun fact, Ervin was AG of Florida until 1964, died in 2004.”

Bluesky · May 13, 2026 · Hallucination

CoCounsel Few voices · 7 of 10

4 voices of 7 criticize it or are mixed about it with a named complaint: hallucination (2), accuracy (2).

“No cocounsel one-shots it that's why it didn't catch the hallucination.”

r/biglaw · Apr 30, 2026 · Hallucination

“More than once Westlaw CoCounsel has claimed a case held a proposition that one side merely asserted in the appeal, even though the judges didn't reach that same conclusion, or may not have reached that issue at all.”

r/Lawyertalk · Jan 9, 2026 · Accuracy

Harvey Few voices · 6 of 10

5 voices of 6 criticize it or are mixed about it with a named complaint: hallucination (4), trust, liability (1).

“All Harvey would need to do is query LexisNexis' system to see if those citations' exist in the database, and if they don't, refresh the generator.”

r/legaltech · Apr 10, 2026 · Hallucination

Perplexity Few voices · 3 of 10

2 voices of 3 criticize it or are mixed about it with a named complaint: hallucination (1), accuracy (1).

“It found a problem in the citation but then hallucinated a different citation to use that does not match at all.”

r/Lawyertalk · Jun 17, 2024 · Hallucination

“Even Perplexity that gives links to the sources, still gives citation wrong when I compare with the relating Rule.”

r/legaltech · Dec 27, 2025 · Accuracy

Copilot Few voices · 2 of 10

1 voice of 2 criticize it or are mixed about it with a named complaint: hallucination (1).

No quote on record that passes our quote rules (verbatim, a statement, no question).

Legora Few voices · 1 of 10

No named complaint on record among its 1 voice.

Spellbook Few voices · 1 of 10

No named complaint on record among its 1 voice.

vLex Vincent Few voices · 1 of 10

1 voice of 1 criticize it or are mixed about it with a named complaint: hallucination (1).

No quote on record that passes our quote rules (verbatim, a statement, no question).

Outside tests, next to the record not our score

Public tests that include citation checking work, with their own publisher, date and caveat. They measure question sets, not lawyers' experience, and are never added to the ranking above. Others are on the arena.

Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools (Magesh et al.)

published 2024-05-30 (arXiv v1; HAI news post dated 2024-05-23) • read 2026-10-22

The preregistered study found Lexis+ AI and Thomson Reuters' Westlaw AI-Assisted Research and Ask Practical Law AI hallucinated 17% to 33% of the time per the arXiv abstract, fewer than GPT-4. The HAI post puts Lexis+ AI and Ask Practical Law AI at more than 17% and Westlaw AI-Assisted Research at more than 34%.

Caveat. Tested 2024 versions; Westlaw range differs between abstract (up to 33%) and HAI post (above 34%); it covered only LexisNexis and Thomson Reuters products plus GPT-4 (no Harvey, Legora or Spellbook), and it ranks nothing overall.

Names: Lexis+ AI, Westlaw AI-Assisted Research, Ask Practical Law AI, GPT-4.

<https://arxiv.org/abs/2405.20362>

Method and cutoff

- **What a voice is.** One public post by a lawyer or legal professional that states an experience or a verdict about one AI tool, read by a language model with one fixed schema (tool, task, sentiment, main complaint). Posts that only ask a question or repeat someone else are not counted. Sources: Reddit, Hacker News, Bluesky, Lemmy, Mastodon, Stack Exchange; posts since Jan 1, 2023.
- **Net and band.** Net = (like – dislike) per 100 voices. The 90% band comes from resampling the same voices 4,000 times; a tool is rated from 10 voices and ordered by the floor of its band, so a few happy voices cannot outrank hundreds. Confidence follows the number of voices: Very low, Low from 10, Medium from 60, High from 300.
- **Few voices.** A tool with fewer than 10 voices on citation checking is listed with its count and no band. It is not a negative finding.
- **Complaints.** Voices whose sentiment is dislike or mixed and that name one main complaint, counted per complaint. The quotes are verbatim, with their source and date.
- **Cutoff.** Data through Oct 26, 2026; the latest voice on citation checking is from Oct 19, 2026. The report is rebuilt with the daily refresh, so the date moves every day.
- **Not in the numbers.** Outside tests are shown apart and never mixed with our figures. Usage is not quality: a tool people use more collects more posts of every

kind. Public posts are not a survey of all lawyers. Independence policy: no vendor pays for a place or a number. Full method: how we count.

Use these numbers: the counts, rates and bands in this report are free to reuse under CC BY 4.0. Credit “aiforlawyers.app (AI Vortex)” with the date and a link to https://aiforlawyers.app/report/citation-checking . Quotes belong to the people who wrote them: quote them with the link to the original post.	Report generated from the public record at https://aiforlawyers.app , cutoff Oct 26, 2026. Numbers are aggregates.
---	---