

# Legal research report, cutoff Oct 26, 2026

As of Oct 26, 2026, aiforlawyers.app counts 1621 lawyer voices on legal research; 1424 voices name one of 14 tools, 11 of which have 10+ voices and are rated. Highest floor of the 90% band: Perplexity (net +88, band +62 to +94, 17 voices, Low confidence).

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

CITE THIS

aiforlawyers.app (AI Vortex), Legal research report, cutoff Oct 26, 2026, <https://aiforlawyers.app/report/legal-research>, licensed CC BY 4.0.

For one figure: “As of Oct 26, 2026, aiforlawyers.app counts 1621 lawyer voices on legal research; 1424 voices name one of 14 tools, 11 of which have 10+ voices and are rated. Highest floor of the 90% band: Perplexity (net +88, band +62 to +94, 17 voices, Low confidence).”

Every tool, with its voices

14 tools named in 1621 voices on legal research. 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  | Perplexity   | Rated      | 17     | +88 | +62 to +94                     | Low        | Accuracy (1)           |
| 2  | Clio         | Rated      | 11     | +64 | +30 to +80                     | Low        | Accuracy (2)           |
| 3  | Claude       | Rated      | 82     | +41 | +27 to +54                     | Medium     | Accuracy (14)          |
| 4  | vLex Vincent | Rated      | 21     | +57 | +26 to +76                     | Low        | Accuracy (3)           |
| 5  | Gemini       | Rated      | 31     | +42 | +19 to +59                     | Low        | Hallucination (6)      |
| 6  | CoCounsel    | Rated      | 168    | +17 | +6 to +27                      | Medium     | Accuracy (41)          |
| 7  | Westlaw AI   | Rated      | 210    | +13 | +4 to +22                      | Medium     | Accuracy (76)          |
| 8  | Lexis+ AI    | Rated      | 184    | -10 | -21 to +0                      | Medium     | Accuracy (80)          |
| 9  | Harvey       | Rated      | 56     | -11 | -27 to +6                      | Low        | Hallucination (11)     |
| 10 | ChatGPT      | Rated      | 623    | -34 | -39 to -28                     | High       | Hallucination (297)    |
| 11 | Copilot      | Rated      | 10     | -30 | -66 to +18                     | Low        | Hallucination (5)      |
| -  | Paxton       | Few voices | 6      | -   | no band: 6 of 10 voices needed | not rated  | Accuracy (1)           |
| -  | Legora       | Few voices | 4      | -   | no band: 4 of 10 voices needed | not rated  | Accuracy (2)           |

| # | Tool      | Status     | Voices | Net | 90% band                       | Confidence | Top complaint (voices) |
|---|-----------|------------|--------|-----|--------------------------------|------------|------------------------|
| - | Spellbook | Few voices | 1      | -   | no band: 1 of 10 voices needed | not rated  | none on record         |

## 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.

### Perplexity Rated · 17 voices

1 voice of 17 criticize it or are mixed about it with a named complaint: accuracy (1).

“I use it too and thinks it’s generally the best LLM, but I’m not confident the sources they pull from are extensive enough for case law research.”

r/Lawyertalk · Feb 12, 2026 · Accuracy

### Clio Rated · 11 voices

3 voices of 11 criticize it or are mixed about it with a named complaint: accuracy (2), price (1).

“I can type my research question and the AI finds the cases and statutes faster than any search I have ever used.”

r/Lawyertalk · May 7, 2026 · Accuracy

“Overall it’s not worth the cost but is adequate for research if you don’t have another platform.”

r/LawFirm · Sep 8, 2026 · Price

### Claude Rated · 82 voices

28 voices of 82 criticize it or are mixed about it with a named complaint: accuracy (14), hallucination (7), trust, liability (5), integration (2).

“I’m still not convinced they are actually saving me time -- just too many errors, too confidently stated, and I have to re-do all the research myself.”

Hacker News · Oct 12, 2026 · Accuracy

“Hey, none of your replies are showing up in my mentions for some reason. But, I mean, we have a motion for sanctions pending against an attorney who used Claude for research and filed a brief full of hallucinated cites back in June, so.”

Bluesky · Sep 23, 2026 · Hallucination

---

## vLex Vincent Rated · 21 voices

5 voices of 21 criticize it or are mixed about it with a named complaint: accuracy (3), hallucination (1), price (1).

*“AI doing research for me has, to date, been almost entirely useless. And that's with vLex.”*

r/legaltech · Dec 21, 2025 · Accuracy

*“Vlex just produced a hallucination. You want something that works and you can trust them just spend the money on the big guys.”*

r/legaltech · Feb 8, 2026 · Hallucination

---

## Gemini Rated · 31 voices

13 voices of 31 criticize it or are mixed about it with a named complaint: hallucination (6), accuracy (5), trust, liability (1), privacy (1).

*“Earlier this year I popped an opposing counsel's ass who filed a document that had 10 case hallucinations out of forty total citations.”*

r/legaltech · Aug 13, 2026 · Hallucination

*“Hey, Gemini as legal research starter smart for spotting gaps, but case law's hit-or-miss.”*

r/Lawyertalk · Oct 5, 2025 · Accuracy

---

## CoCounsel Rated · 168 voices

79 voices of 168 criticize it or are mixed about it with a named complaint: accuracy (41), hallucination (16), price (16), trust, liability (3), integration (3).

*“I wiped his ass with three hour long responses that actually cited controlling rules, because co counsel didn't give a damn about several small nuances it deemed irrelevant.”*

r/Lawyertalk · Aug 2, 2026 · Accuracy

*“It straight up has hallucinated holdings and has informed me that statutes I know exist don't.”*

r/legaltech · Aug 29, 2026 · Hallucination

---

## Westlaw AI Rated · 210 voices

109 voices of 210 criticize it or are mixed about it with a named complaint: accuracy (76), hallucination (18), price (10), trust, liability (5).

*“I cannot wait for someone to tear down Westlaw's walled garden. Their AI sucks so bad”*

r/biglaw · Sep 18, 2026 · Accuracy

*“It produces some great work product if you don't mind that how confidently wrong it is 50% of the time.”*

r/biglaw · Oct 9, 2026 · Hallucination

---

## Lexis+ AI Rated · 184 voices

112 voices of 184 criticize it or are mixed about it with a named complaint: accuracy (80), hallucination (19), price (9), trust, liability (3), integration (1).

*"I don't find it very effective for drafting or anything else really. Even the research it gives me is subpar."*

r/Lawyertalk · Sep 17, 2026 · Accuracy

*"(I am ethically opposed to AI but Protege hallucinates like crazy)."*

r/Lawyertalk · Jul 5, 2026 · Hallucination

---

## Harvey Rated · 56 voices

28 voices of 56 criticize it or are mixed about it with a named complaint: hallucination (11), accuracy (9), price (6), trust, liability (2).

*"It's not long ago (12mo max) that they were guiding people away from using it as a legal research tool in their official training (presumably because of the hallucination risk, which they can never fully eliminate)."*

r/legaltech · Jul 20, 2026 · Hallucination

*"Harvey AI only gets some content from Lexis. They have the Lexis Ask pulling from US Primary Law with Shepard's. No secondary material."*

r/legaltech · Feb 2, 2026 · Accuracy

---

## ChatGPT Rated · 623 voices

452 voices of 623 criticize it or are mixed about it with a named complaint: hallucination (297), accuracy (119), trust, liability (27), privacy (6), integration (3).

*"Imagine hiring a veteran lawyer with 30+ years of experience, only for him to submit 6 completely fabricated court cases in federal court because he trusted ChatGPT..."*

Bluesky · Oct 16, 2026 · Hallucination

*"I've definitely had to do this for tech start-up CEOs and literally have ChatGPT, Claude, etc. admit to them why they initially dispensed completely wrong previous advice/analysis."*

r/Lawyertalk · Sep 12, 2026 · Accuracy

---

## Copilot Rated · 10 voices

7 voices of 10 criticize it or are mixed about it with a named complaint: hallucination (5), accuracy (1), trust, liability (1).

*"The second was EAT, also didn't deal with the point and had a citation that was out by 17 years."*

Bluesky · Sep 15, 2026 · Hallucination

*"They've all been wrong to a greater or lesser extent."*

Bluesky · Sep 24, 2024 · Accuracy

---

---

## Paxton Few voices · 6 of 10

1 voice of 6 criticize it or are mixed about it with a named complaint: accuracy (1).

“It actually is mostly accurate although sometimes case citations are wonky / not correct/applicable”

r/Lawyertalk · Oct 28, 2025 · Accuracy

---

## Legora Few voices · 4 of 10

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

“At least where I reside, Legora cannot be trusted to do good/accurate research.”

r/legaltech · Oct 16, 2026 · Accuracy

---

## Spellbook Few voices · 1 of 10

No named complaint on record among its 1 voice.

---

## Outside tests,

next to the  
record (not our  
score)

Public tests that include legal research 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.

LEGAL-SPECIFIC VENDOR TEST · VALS AI

### Vals Legal AI Report: Legal Research (VLAIR Legal Research)

published 2025-10-14 · read 2026-10-22

Four products (Alexi, Counsel Stack, Midpage, ChatGPT) and a lawyer baseline answered 200 US legal research questions. Weighted scores were 74% to 78% for the AI products and 69% for the lawyers; the page says Counsel Stack received the highest score across all criteria, and that legal AI and generalist AI both scored about 80% on accuracy (78 to 81% legal AI, 80% ChatGPT).

**Caveat.** Only four products took part; the page lists Reed Smith, Fisher Philips, McDermott, Ogletree, Paul Hastings and Paul Weiss as the question-writing consortium. Trade press (legaltechnology.com headline, not re-read here) says the market leaders were absent. Lawyer baseline answers were shorter by default, which the report says lowered their scores.

Names: Alexi, Counsel Stack, Midpage, ChatGPT, Lawyer Baseline.

<https://www.vals.ai/industry-reports/vlair-10-14-25>

### **Vals Legal Research Bench (agentic US legal research leaderboard)**

published 2026-10-22 (page shows Updated 10/22/2026) · read 2026-10-22

As of October 22, 2026 the page reports Muse Spark 1.3 Max, Claude Opus 5 and Claude Fable 5.1 tied for first at 55.29% (all-pass accuracy). Next: Gemini 4 Argon 54.81%, Wonder Full 53.37%, Wonder and Claude Opus 5.5 both 50.48%, GPT-5.6 Sol 48.08%.

**Caveat.** Tests raw models in an agent harness, not legal products like Harvey or CoCounsel; private test set; 'All-Pass' means every rubric item must pass. Top 3 are a three-way tie, so rank order is arbitrary.

Names: Muse Spark 1.3 Max, Claude Opus 5, Claude Fable 5.1, Gemini 4 Argon, Wonder Full, Wonder.

[https://www.vals.ai/benchmarks/legal\\_research](https://www.vals.ai/benchmarks/legal_research)

### **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>

### **LegalBench-RAG (Pipitone and Houir Alami)**

published 2024-08-19 · read 2026-10-22

Introduced as the first benchmark designed to evaluate the retrieval step of RAG pipelines in the legal domain, built by tracing LegalBench query contexts back to their original locations in a legal corpus, and emphasizing minimal, precise snippets. The abstract gives no model scores.

**Caveat.** Measures retrieval only, not answer quality; no scores in the abstract; 2024 paper.

Names: RAG retrieval pipelines (no products named in the abstract).

<https://arxiv.org/abs/2408.10343>

**Vals CaseLaw (Canadian court case Q&A, archived)**

published 2026-05-04 (page shows Updated 5/4/2026) · read 2026-10-22

Archived because performance saturated and it is no longer run on new models.  
Final board: 1 Grok 4.3 79.31%, 2 GPT 5.1 73.42%, 3 GPT 4.1 69.88%; Claude Opus 4.7 68.38%.

**Caveat.** Canadian case law only, private questions, and the page says it is archived, so it says nothing about 2026-09 and later models.

Names: Grok 4.3, GPT 5.1, GPT 4.1, GPT 5 Mini, Claude Opus 4.7.

[https://www.vals.ai/benchmarks/case\\_law\\_v2](https://www.vals.ai/benchmarks/case_law_v2)

## 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 legal research 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 legal research is from Oct 21, 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/legal-research>. 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.