

Contract review report, cutoff Oct 26, 2026

As of Oct 26, 2026, aiforlawyers.app counts 305 lawyer voices on contract review; 248 voices name one of 10 tools, 4 of which have 10+ voices and are rated. Highest floor of the 90% band: Claude (net +46, band +29 to +59, 65 voices, Medium confidence).

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

CITE THIS

aiforlawyers.app (AI Vortex), Contract review report, cutoff Oct 26, 2026,
<https://aiforlawyers.app/report/contract-review>, licensed CC BY 4.0.

For one figure: “As of Oct 26, 2026, aiforlawyers.app counts 305 lawyer voices on contract review; 248 voices name one of 10 tools, 4 of which have 10+ voices and are rated. Highest floor of the 90% band: Claude (net +46, band +29 to +59, 65 voices, Medium confidence).”

Every tool, with its voices

10 tools named in 305 voices on contract review. 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	65	+46	+29 to +59	Medium	Accuracy (12)
2	Spellbook	Rated	51	+39	+20 to +55	Low	Accuracy (9)
3	Harvey	Rated	43	+12	-7 to +30	Low	Accuracy (17)
4	ChatGPT	Rated	57	-11	-29 to +9	Low	Accuracy (22)
-	Copilot	Few voices	9	-	no band: 9 of 10 voices needed	not rated	Accuracy (3)
-	CoCounsel	Few voices	6	-	no band: 6 of 10 voices needed	not rated	Accuracy (3)
-	Legora	Few voices	6	-	no band: 6 of 10 voices needed	not rated	Accuracy (2)
-	Luminance	Few voices	6	-	no band: 6 of 10 voices needed	not rated	Accuracy (4)
-	Gemini	Few voices	3	-	no band: 3 of 10 voices needed	not rated	none on record
-	Perplexity	Few voices	2	-	no band: 2 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.

Claude Rated · 65 voices

18 voices of 65 criticize it or are mixed about it with a named complaint: accuracy (12), privacy (2), integration (2), hallucination (1), trust, liability (1).

“Explaining to people why Claude’s positions are unreasonable has become a part of my day to day job.”

r/Lawyertalk · Jul 31, 2026 · Accuracy

“However using claude for building own contract analyzer , won't give you data privacy as you would be using claude anyway.”

r/legaltech · May 9, 2026 · Privacy

Spellbook Rated · 51 voices

19 voices of 51 criticize it or are mixed about it with a named complaint: accuracy (9), price (5), hallucination (3), trust, liability (1), privacy (1).

“Out of say, 30 comments it had in analyzing an agreement, most seemed like noise (possibly technically valid, but not helpful) and one or two were interesting and useful but far from being worth the fee.”

r/legaltech · Sep 19, 2026 · Accuracy

“I don’t mind useful wraps, but the price point in this one is so steep...”

r/legaltech · Mar 6, 2026 · Price

Harvey Rated · 43 voices

24 voices of 43 criticize it or are mixed about it with a named complaint: accuracy (17), hallucination (2), trust, liability (2), price (2), integration (1).

“I use Harvey, Claude, etc. for diligence and know for a fact it is still leagues behind in picking up change of control/anti-assignment provisions.”

r/biglaw · Mar 13, 2026 · Accuracy

“I wish I could upload my back and forth with Harvey today where it incorrectly cited sections over and over again and made up signatories in a checklist.”

r/biglaw · Mar 13, 2026 · Hallucination

ChatGPT Rated · 57 voices

32 voices of 57 criticize it or are mixed about it with a named complaint: accuracy (22), hallucination (5), trust, liability (2), privacy (2), integration (1).

“I find that when I feed contracts to ChatGPT (which I do all the time) it will try to find 50 things wrong with it.”

Hacker News · Sep 18, 2026 · Accuracy

“I uploaded a contract and asked it a question and it started citing nonexistent sections.”

r/biglaw · Nov 25, 2025 · Hallucination

Copilot Few voices · 9 of 10

3 voices of 9 criticize it or are mixed about it with a named complaint: accuracy (3).

“Even procurement is running contracts, order forms etc through whatever GenAI just to produce completely unusable risk lists.”

r/Lawyertalk · Aug 14, 2026 · Accuracy

CoCounsel Few voices · 6 of 10

4 voices of 6 criticize it or are mixed about it with a named complaint: accuracy (3), trust, liability (1).

“My CoCounsel demo was a disaster. Spellbook and its ilk are good.”

r/Lawyertalk · Nov 1, 2024 · Accuracy

“I’ve used AI stuff like Spellbook, CoCounsel, AI Lawyer, and the pattern I keep seeing is that the analysis label doesn’t matter if the UI encourages people to accept changes like autopilot.”

r/legaltech · Dec 26, 2025 · Trust, liability

Legora Few voices · 6 of 10

4 voices of 6 criticize it or are mixed about it with a named complaint: accuracy (2), hallucination (1), price (1).

“Legora and Harvey price most law firms and even some in house teams out, but in my trials, they don’t add any value on the contract analysis side and the results of the suggested redlines are better with other tools.”

r/legaltech · Nov 13, 2025 · Accuracy

“We also tried Legora’s tabular review, there were also numerous hallucinated provisions in our due diligence exercise.”

Hacker News · Sep 18, 2026 · Hallucination

Luminance Few voices · 6 of 10

4 voices of 6 criticize it or are mixed about it with a named complaint: accuracy (4).

“My comments about how Luminance didn't work at all for me have been downvoted multiple times and there's a four year old account with zero previous comments coming to life and saying Spellbook is great.”

r/legaltech · Dec 19, 2024 · Accuracy

Gemini Few voices · 3 of 10

No named complaint on record among its 3 voices.

Perplexity Few voices · 2 of 10

No named complaint on record among its 2 voices.

Outside tests,

next to the
record not our
score

Public tests that include contract review 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 (VLAIR): Feb 2025 edition, legal tasks

published 2025-02-27 · read 2026-10-22

Four tools were tested on seven tasks against a lawyer baseline: CoCounsel, Vincent AI, Harvey Assistant and Oliver (Lexis+ AI withdrew from the sections studied). Harvey took the top score on five of the six tasks it entered, including Document Q&A at 94.8% and Chronology Generation at 80.2%; CoCounsel scored 89.6% on Document Q&A and 77.2% on Document Summarization, the top score there. The lawyer baseline beat the AI tools on EDGAR research and on Redlining (79.7%).

Caveat. Vendors opted in task by task (Harvey six of seven, CoCounsel four), so tools were not compared on the same tasks; 2025 versions, over a year old; not a ranking of all legal AI.

Names: Harvey Assistant, CoCounsel, Vincent AI (vLex), Oliver (Vecflow, now August), Lexis+ AI (withdrew).

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

Harvey Legal Agent Benchmark (LAB) leaderboard, hosted by Mercor

published 2026-05-06 (launch per Harvey blog coverage); leaderboard date not shown on page · read 2026-10-22

Open-source benchmark of 1,749 tasks across 24 practice areas plus in-house contracting, graded with all-pass rubrics (100,000+ criteria). Pass@1 top 5 on the page: Gemini 4 Argon High 27.4% (± 5.4), Grok 4.7 xHigh 19.5% (± 4.5), Muse Spark 1.2 xHigh 19.4% (± 4.6), Claude Opus 5 Max 17.8% (± 1.5), Claude Fable 5.1 Max 17.4% (± 4.2).

Caveat. Built by Harvey (a vendor); the leaderboard lists foundation models, not Harvey's own product; leaderboard publication date not shown and the top-5 error bars overlap heavily; the leaderboard page counts 1,749 tasks versus 'more than 1,200' reported at launch, so the set has grown.

Names: Gemini 4 Argon, Grok 4.7, Muse Spark 1.2, Claude Opus 5, Claude Fable 5.1.

<https://www.mercor.com/apex/partner-benchmarks/harvey-lab-leaderboard/>

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 contract review 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 contract review is from Oct 16, 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/contract-review>. 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.