Methodology

Z-Labs, LLC · Effective April 15, 2026 · v1.2

Live version: https://hallucinx.com/methodology

---

1. What HallucinX does

HallucinX verifies that the case citations in a legal brief actually exist and that the case name and metadata in the brief match the opinion they purport to cite. It does not evaluate the legal argument. It does not check whether a case is still good law. It does not catch substantive errors in legal reasoning. It catches one specific class of error: a citation that doesn't resolve to the case the brief names — either the reporter coordinates resolve to no case at all, or they resolve to a different case than the one named.

The verification is deterministic. Every classification has a concrete reason an attorney can show a judge. There is no language model of any kind in the verification path — no large language model (LLM), no generative AI, no statistical text model.

2. How it works

The provenance chain runs end to end as follows. An attorney walking a judge through the tool's behavior should be able to point at each step.

1. The brief is uploaded into the user's browser as a PDF.
2. Citation strings are extracted in-browser using *eyecite*, an open-source citation-parsing library originally developed by Free Law Project. The original document never leaves the browser for verification — only the extracted citation strings.
3. The citation strings are sent to CourtListener's Citation Lookup API. CourtListener is operated by Free Law Project, a 501(c)(3) nonprofit, and its database contains more than ten million judicial opinions across federal and state jurisdictions.
4. For each citation, CourtListener returns the matching opinion if one exists. HallucinX then compares the case name and metadata in the brief against what CourtListener returned. Disagreements feed the fabrication heuristics in Section 4.
5. Each citation is assigned one of five states based on what was found.

3. The five states

Every citation in a HallucinX report carries one of five classifications. The classifications are exclusive — a citation lands in exactly one state — and ordered below by how much attention they require.

Verified — the citation exists and the metadata matches. This is the state where automatic verification has succeeded. Spot-checking is still part of the attorney's duty of competence; the other four states call for closer follow-up because automatic verification fell short or did not apply.

Caution — the citation exists but a discrepancy was detected. Common causes are case-name drift between the brief and the indexed opinion or a year mismatch. Caution is most often a real case with a clerical error in the brief; sometimes it is the early sign of a fabrication. Review before relying on the cite.

Check manually — the citation could not be verified automatically, and the cited reporter or court appears to be outside CourtListener's coverage. State courts and unpublished opinions land here most often. This is not a fabrication signal. Confirm the citation by another means — Westlaw, Lexis, the issuing court, or the official reporter.

Unverified — no match was found in CourtListener and no fabrication signature matched. The citation could be a real case in a coverage gap, or a fabrication that does not match a known pattern. Confirm by another means before relying on the cite.

Alert — high-confidence fabrication. One or more deterministic fabrication heuristics fired. The next section enumerates them. An Alert classification means CourtListener's response is structurally inconsistent with the citation being a real case; treat as fabrication unless and until proven otherwise.

4. The fabrication heuristics

When CourtListener returns no match, or returns a match that disagrees with the brief, HallucinX applies a small number of deterministic heuristics to decide whether the result is a coverage gap or a fabrication. Each heuristic has a code (H1, H2, and so on) that appears in the report's reasons column for any citation it fires on. Each is the result of corpus analysis against known fabrications and known-good briefs, and each carries a known false-positive risk that an attorney reviewing a flag can weigh.

Page 2 of 8

H1 — Reporter points to a different case

Pattern. CourtListener resolves the citation (HTTP 200), but the case name returned by CourtListener and the case name as it appears in the brief share zero meaningful words. "Meaningful" means longer than two letters and not a common stopword (Inc, Corp, the, of, in, and, etc.). Pincite-form short citations and string-citation extraction failures are excluded.

Why this is a fabrication signal. A real citation cannot disagree with the reporter on every meaningful word of the case name. The brief is asserting that volume, reporter, and page point at a named case; the reporter's database says they point at an entirely different one. The classic AI tell is an invented case name pasted onto a real reporter citation that belongs to an unrelated case.

Example. In *Mata v. Avianca*, the brief cited "Peterson v. Iran Air, 905 F. Supp. 2d 121 (D.D.C. 2013)." That reporter location does resolve in CourtListener — but to a different case (an ISS Marine Services matter), with no meaningful word in common with "Peterson" or "Iran Air."

False-positive risk. Low. Real briefs do not contain full case-name disagreements with the reporter's index.

H2 — Federal reporter coverage is complete; a 404 is a fabrication signal

Pattern. CourtListener returns 404 for a citation whose reporter is in the complete-coverage set — U.S., S. Ct., L. Ed. and L. Ed. 2d, F. and F.2d/3d/4th, F. Supp. and F. Supp. 2d/3d, B.R., F.R.D. — and the cited year is more than six years before today. Government-prosecution captions ("United States v. X," "People v. X," "State v. X," "Commonwealth v. X," municipal actions) and pincite-form short cites are excluded.

Why this is a fabrication signal. CourtListener's coverage of these published federal reporters is effectively complete for opinions older than about six years. Recent volumes have indexing lag, which is why the heuristic excludes them. A 404 on a 2017 F.3d citation is not a coverage gap; the opinion would be in the database if it existed.

Example. In *Mata*, the brief cited "Varghese v. China Southern Airlines Co. Ltd., 925 F.3d 1339 (11th Cir. 2019)." The citation 404s. F.3d is in the complete-coverage set, and 2019 is more than six years before today.

False-positive risk. Low.

H3 — Unpublished state-court citation not found

Pattern. CourtListener returns 404 for a citation formatted as a non-precedential state-court disposition. Recognized formats include Illinois Rule-23 unpublished orders ("YYYY IL App (xxxx) NNNNN-U"), Illinois supreme-court unpublished ("YYYY IL NNNNN-U"), New York Slip Op unpublished ("YYYY NY Slip Op NNNNN(U)"), state Unpub. LEXIS, Ohio App. Unpub., and Texas App. unpub. variants.

Why this is a fabrication signal. Unpublished state dispositions are the formats AI fabrications most often mimic, because they are the hardest to verify by hand and are non-precedential, which reduces the chance a reader will check. CourtListener's coverage of unpublished state opinions is genuinely thin, so a 404 alone is not decisive. Combined with the format itself — which is what the heuristic gates on — it becomes a meaningful flag.

Example. In *Mata*, the brief cited "Shaboon v. EgyptAir Airlines Co., 2013 IL App (1st) 111279-U." The citation matches the Illinois -U format and 404s.

False-positive risk. Medium. CourtListener does have real coverage gaps in unpublished state opinions. Treat an Alert classification backed only by H3 as a strong prompt to verify by another means rather than as a finding of fabrication on its own.

H4 — State-appellate Westlaw citation not found

Pattern. CourtListener returns 404 for a Westlaw citation ("YYYY WL NNNNNNN") in state intermediate-appellate context — a parenthetical naming a state appellate court (Tex. App., Ga. Ct. App., Ill. App. Ct., App. Div., and similar) — where the cited WL year is at least two years before today and a brief-side case name is present.

Why this is a fabrication signal. Westlaw indexing for state intermediate-appellate opinions older than two years is generally complete in CourtListener. A 404 in this context, on a plausibly-formed case name, is more consistent with fabrication than with a coverage gap.

Example. In *Mata*, two such citations appeared: "Martinez v. Delta Air Lines, Inc., 2019 WL 4639462 (Tex. App. Sept. 25, 2019)" and "Estate of Durden v. KLM Royal Dutch Airlines, 2017 WL 2418825 (Ga. Ct. App. June 5, 2017)." Both 404, both carry state-appellate parentheticals, and both are old enough to be indexed.

False-positive risk. Medium-high. State Westlaw coverage in CourtListener is the least uniform of the heuristic's inputs. Treat as a strong prompt for manual verification.

H5 — Westlaw number outside the plausible range for its year

Pattern. The Westlaw sequence number in a citation ("YYYY WL NNNNNNN") exceeds an empirical per-year ceiling derived from Westlaw's published indexing volume, with roughly thirty percent headroom above the observed annual maximum. This heuristic fires regardless of CourtListener's response — the WL number is arithmetically impossible whether CourtListener returns 200, 404, or anything else.

Why this is a fabrication signal. Westlaw assigns sequence numbers monotonically within each year. An impossibly high number cannot have been issued; there were not that many opinions indexed that year. The number is wrong on its face.

Example. A fabricated cite of "2019 WL 11199934" carries an 11.2-million WL sequence number for 2019; the empirical upper bound for 2019 is approximately eight million, padded to a ten-million ceiling for safety.

False-positive risk. Very low. The ceiling is set well above any observed real-world WL number.

Across all five heuristics, HallucinX was tuned conservatively. Gating conditions are tight, and a citation in a genuine coverage gap is treated as Check manually or Unverified rather than Alert. In benchmark testing against 774 citations drawn from a clean corpus of recent federal-circuit opinions, approximately 10.85% of legitimate citations were classified as Unverified or Alert and required attorney follow-up. Opinions are a stricter test bed than briefs — opinions string-cite and short-form-cite in ways that introduce extraction noise — so briefs, which are the tool's intended input, are projected to flag at a lower rate; that briefs-specific rate has not yet been independently benchmarked. On the *Mata v. Avianca* corpus, all six fabricated citations classify as Alert.

5. Known limitations

HallucinX's confidence comes from a narrow scope. The list below is the full inventory of what the tool does not do and what it does not cover. An attorney defending the use of the tool to a judge should be able to point at this section as the boundary of the tool's claims.

Coverage limitations

- State-court coverage in CourtListener is uneven. Some jurisdictions are nearly complete; others are sparse. A "Check manually" classification is particularly common for state appellate opinions outside the coverage set.
- Unpublished opinions are inconsistently indexed across all jurisdictions.

- Recent reporter volumes — typically those issued within the last six to twelve months — may not yet be in CourtListener.

What the tool does not do

- HallucinX does not check whether a case is still good law. It is not a substitute for Shepardizing or KeyCite.
- It does not evaluate substantive legal reasoning. A brief can have all-verified citations and still be wrong on the law.
- It does not catch quote-misattribution where a quote does exist in the cited opinion but is being used out of context.
- It does not verify pinpoint citations to specific page numbers within an opinion.
- It does not verify that parallel citations point to the same case, though it does verify each citation independently.

What the tool may miss

HallucinX verifies citations the extractor finds. Citations the extractor fails to find pass through unverified — and silently. The extraction layer (eyecite, applied to the text the browser pulls out of the PDF) is robust on standard typography but can miss citations distorted by:

- **OCR errors in scanned briefs.** A scanned PDF whose OCR misread a digit, a comma, or a reporter abbreviation can yield a citation string the parser does not recognize.
- **Unusual whitespace and line breaks.** Non-breaking spaces, tabs, soft hyphens, or citations broken across line, column, or page boundaries can prevent the extractor from assembling a complete citation string.
- **Non-standard typography.** Smart quotes, ligatures, Unicode dashes other than the standard hyphen-minus, and unusual font substitutions can produce citation strings the parser does not match.

The practical mitigation is a citation-count check. After running a brief through the tool, count the citations in the brief by hand or by skim and compare to the count HallucinX reports. A mismatch is worth investigating: it may mean a citation was missed by the extractor and is not represented in the report at all. HallucinX does not claim a coverage guarantee on what the extractor sees; an attorney's own count is the check on what it does not.

What attorneys are expected to do

- Review every classification. Caution, Check manually, and Unverified require closer attention; Verified citations are not exempt from spot-checking.
- Treat the tool as an audit aid, not a final verification.
- Comply with their jurisdiction's professional responsibility rules regardless of tool output. HallucinX does not, and cannot, discharge those duties for the attorney who files the brief.

6. Why no language model in the verification path

The most common form of citation hallucination is a language model producing a citation string that does not correspond to a real case. Using another language model to verify the first one's output reproduces the same failure mode at a different layer. Probabilistic verification of probabilistic generation is not verification.

HallucinX uses deterministic comparison against an external authoritative database. When the tool says a citation is fabricated, it means CourtListener returned no match for the volume and page, or returned a case whose name shares zero meaningful words with what the brief claimed. When the tool says a citation is verified, it means CourtListener returned an opinion at that reporter location whose case name and metadata match what the brief claims. Every classification has a concrete reason that can be reproduced by anyone with access to the same public database.

A judge does not have to trust the tool's judgment. The judge can verify the tool's judgment.

7. Open source and reproducibility

Verification by HallucinX rests on artifacts a judge or opposing counsel can independently inspect.

- **CourtListener is public.** Anyone can independently look up any citation HallucinX classifies and confirm what CourtListener returns.
- **The eyecite citation-parsing library is open source.** The extraction logic that turns brief text into citation strings is inspectable code, not a proprietary black box.
- **The same parser runs twice.** eyecite runs once in the user's browser to extract citations from the brief, and again on CourtListener's server against the strings it receives. Both passes converge on the same canonical citation form, so the lookup key the brief generates and the lookup key CourtListener resolves against are produced by the same open-source code.

- **The fabrication heuristics are deterministic and documented.** Section 4 above is the full set; their gating conditions are described there in plain English and live in the engine source as straightforward code.
- **The citation strings sent to CourtListener are not confidential client information.** Under ABA Model Rule 1.6 and its state analogues, public references — citations to published judicial opinions — are not client confidences. HallucinX does not transmit document contents, party names, or any client-identifying detail to CourtListener; only the citation strings themselves. (See Ethics and use for the bar association guidance on this point.)
- **HallucinX retains no documents.** Briefs are parsed in the user's browser. For verification, the original document never reaches HallucinX's servers — only the extracted citation strings do. For the optional annotated-brief download, the original PDF transits the server in memory and is returned annotated; it is not written to disk, logged, or stored at any layer. No log of brief content is kept in either flow. (See Privacy for details.)

8. Contact

Methodology questions from attorneys, judges, or opposing counsel can be directed to contact@hallucinx.com. We do not promise a particular response time.

Nothing on this page constitutes legal advice. Attorneys remain responsible for verifying citations and complying with their jurisdiction's professional responsibility rules regardless of the output of this or any tool. Spot-check every citation. HallucinX is built to make that workflow efficient, not to replace it.