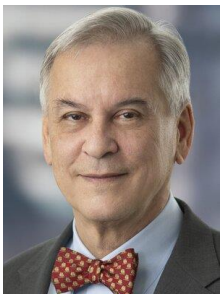
How Judiciary Can Minimize AI Risks In Secondary Sources

By **John Siffert and Allison Morse** (October 31, 2024)



John Siffert

Allison Morse

The judiciary's current approach to preventing lawyers from citing phantom cases that have been hallucinated by generative artificial intelligence may not go far enough to exorcise all the potential demons unleashed by the new technology.

At the time of writing, over 30 courts have issued standing orders addressing the use of generative AI in briefs, pleadings and other filings. The common feature of these AI orders is a requirement that attorneys declare that the filer reviewed the source material and personally verified its accuracy.[1]

Because AI orders and other safeguards[2] do not extend to secondary source materials, the judiciary should take steps to protect itself against the risk of hallucinations from secondary sources.

In effect, the AI orders seek to discourage reliance on phantom precedent by shifting the burden onto the authors of a filing by requiring them to go beyond the current requirements of Rule 11 of the Federal Rules of Civil Procedure in the case that the content of the filing is the product of using generative AI.[3] Rule 11 essentially requires attorneys to sign court filings, thus verifying that the claims therein are made in good faith and are reasonably supported.

The need for AI orders is underscored by academic research revealing that even the top-of-the-line, legal-specific generative AI research tools continue to hallucinate, at least for the time being.[4]

Nonetheless, the need for AI orders is not universally accepted. Indeed, the U.S. Court of Appeals for the Fifth Circuit declined to issue an AI order in the face of public comments that Rule 11 and other regulations already hold lawyers responsible for assuring the accuracy of their filings.

Regardless of whether current AI orders are needed to reach risks not contemplated by Rule 11, the orders require attorneys to certify only the existence and accuracy of primary authorities that lawyers routinely rely upon, such as judicial opinions, statutes and regulations.

Secondary sources like treatises, however, present a thornier problem.

Some authors and sources — for example, Weinstein's Evidence, or Prosser and Keeton on the Law of Torts — are so widely recognized and revered that their articles and publications carry the weight of authority and have the capacity to influence the outcome of a case or controversy.

Likewise, there are some secondary sources — for example, the various Restatements of the Law — that are viewed as having precedential effect.

Both lawyers and judges rely on these published secondary source materials and take for

granted their trustworthiness, citing these sources directly in their filings and opinions. It is therefore crucial that these materials are accurate; indeed, an inaccuracy in secondary materials could infect a filing submitted to a court in the same way as a hallucinated case in a generative AI-drafted brief.

Westlaw and Lexis have released their own generative AI to assist in legal research. It is possible that they, and other major legal publishers, may soon use AI in furtherance of their own publications. Assuming that is possible, if not probable, there is the very real risk that secondary source materials soon will be — if they have not already been — infected by the hallucinations inherent in generative AI.

Unlike primary source materials, it is difficult, if not impossible, for an attorney personally to verify that a secondary source is free from hallucinations. Unless the publisher discloses the use of generative AI in drafting or updating a secondary source, an attorney referencing it in researching or drafting a brief is unlikely to even be aware of the potential for hallucination.

The AI orders in their current form do not require a certification that the secondary sources are free of phantoms. Indeed, it would be ineffective, if not fundamentally unfair, to shift the burden of verifying the accuracy of a secondary source drafted with the assistance of generative AI to an attorney who relies upon it.

It does not appear that courts are empowered to issue AI orders directed to publishers who are not appearing as counsel in cases before them. That does not mean that courts are without recourse to stop AI hallucinations in secondary materials from creeping into the courtroom.

For example, the board of judges of a district or circuit court could request publishers and database hosts of AI-generated materials to warrant the existence and accuracy of the primary authorities cited within their publications — or to adopt the language of the proposed modification of Rule 901(b)(9) of the Federal Rules of Evidence that the results are valid and reliable.

The judiciary's request to publishers could include the admonition that courts within the jurisdiction will be free to disregard any secondary source that fails to contain a verification that the publisher abides by the tenets of best AI practices.

It is quite difficult to imagine that publishers and database hosts would disregard such a communication from the judiciary. Indeed, it would send a convincing message that publishers and database hosts are expected to be gatekeepers, and that courts recognize that legal publishers and databases are uniquely situated to protect the courts from citations to phantom authority.

This type of certification is particularly well-suited to the secondary source material context. Rule 11 and other existing safeguards that arguably make AI orders redundant do not apply to secondary sources like treatises.

Although it would be voluntary for publishers and database hosts to warrant the existence and accuracy — or validity and reliability — of the contents of their publications, doing so would be an important acknowledgment of their responsibility to prevent AI hallucinations from infecting judicial decisions.[5]

In the face of the risks posed by AI hallucinations, even in the most sophisticated formats, it

would be prudent for the judiciary to enlist legal publishers and database hosts in policing against the unwanted effects of AI in court decisions.

John S. Siffert is a founding partner and Allison Morse is an associate at Lankler Siffert & Wohl LLP.

[1] See, e.g., https://www.documentcloud.org/documents/24747791-blumenfeld_ai-order.

[2] One example of another potential trial safeguard is a contemplated modification of current Fed. R. Evid. 901(b)(9) for AI evidence requiring that AI evidence may be admitted only upon proof that the machine process or system "produces a valid and reliable result."

[3] Fed. R. Civ. P. 11 provides that the signatures required for "pleadings, motions and other papers" mean that the person signing "certifies that to the best of the person's knowledge, information and belief, formed after an inquiry reasonable under the circumstances:

(1) It is not being presented for any improper purpose…

(2) the claims, defenses and other legal contentions are warranted by existing law …;

(3) the factual contents have evidentiary support…; and

(4) the denials of factual contentions are warranted on the evidence, or…reasonably based on belief…"

[4] In a pre-print study, researchers at Stanford University found that legal research products incorporating generative AI produced by Lexis and Westlaw had hallucination rates north of 17%. See Magesh & Surani et al., Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools (pre-print, 2024). The creators of these tools have maintained that internal benchmark testing reveals much lower rates of hallucinations. See https://www.law.com/legaltechnews/2024/06/04/updated-stanford-report-finds-high-hallucination-rates-on-westlaw-ai/.

[5] See, e.g., https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/developing-ai-responsibly-the-lexisnexis-commitment?srsltid=AfmBOoq-JIovsY9_-8znO4PE9tUZjep29bfL4VTKV7QpQBXacPlt43N5; https://www.thomsonreuters.com/en/artificial-intelligence/ai-principles.html.