

**DIVE BRIEF**

# Legal GenAI tools mislead 17% of time: Stanford study

Specialized legal applications are better than a general purpose tool like ChatGPT but whether they save time, given the oversight they need, is a question, an academic study says.

Published May 24, 2024



Robert Freedman  
Lead Editor

*Stress, deadline and pressure with a young employee in the professional workplace on a dark evening. Jacob Wackerhausen via Getty Images*

**Dive Brief:**

- Purpose-built generative AI legal tools continue to be plagued by hallucinations – 17% of the time – and shouldn’t be used without close scrutiny, casting doubt on how much they save time and effort, a study by Stanford University’s Institute for Human-centered Artificial Intelligence finds.
- “Given the high rate of hallucinations, lawyers may find themselves having to verify each and every proposition and citation provided by these tools, undercutting the stated efficiency gains that legal AI tools are supposed to provide,” the study report says.
- Even so, the tools are a big improvement over a general purpose tool like ChatGPT, which has an error rate as high as 82% when used for legal purposes.

**Dive Insight:**

Companies behind some of these purpose-built legal tools say they've gone a long way to reducing hallucinations by using what's called a retrieval-augmentation generation (RAG) process, in which they integrate a language model with a database of legal documents.

"Relying on RAG, leading legal research services have released AI-powered legal research products that they claim 'avoid' hallucinations and guarantee 'hallucination-free' legal citations," the report says.

But even when using a RAG process, the legal tools can still get the law or citation wrong, the report says.

The citation errors could especially be a problem because they tend to be trickier to identify.

"Given the critical importance of authoritative sources in legal research and writing, the [citation] hallucination may be even more pernicious than the outright invention of legal cases," the report says. "A citation might be 'hallucination-free' in the narrowest sense that the citation exists, but that is not the only thing that matters. The core promise of legal AI is that it can streamline the time-consuming process of identifying relevant legal sources. If a tool provides sources that seem authoritative but are in reality irrelevant or contradictory, users could be misled. They may place undue trust in the tool's output, potentially leading to erroneous legal judgments and conclusions."

To test how well purpose-built tools do, researchers created a dataset of more than 200 open-ended legal queries, in some cases premised on a mistaken understanding of the law to see if the tools could recognize the mistake.

In one example, researchers asked why Supreme Court Associate Justice Ruth Bader Ginsburg dissented in *Obergefell*, a case about

same-sex marriage in which Ginsburg joined. One of the tools gave an answer that was wrong in two ways: It agreed with the mistaken premise that she dissented and also mistook Obergefell as a case about copyright law.

In another example, researchers asked about the standard of review that applies to abortion regulations under the U.S. Constitution. One of the tools gave an answer pointing to the undue burden test in Casey, a standard that was overruled in 2022 by the landmark Dobbs case.

“Legal hallucinations have not been solved,” the researchers say. “The legal profession should turn to public benchmarking and rigorous evaluations” before relying on the tools in their practice.

### **LexisNexis response**

In response to the Stanford study, Jeff Pfeifer, LexisNexis chief product officer for the U.S., Canada, Ireland and U.K., provided the following response. LexisNexis offers one of the purpose-built legal tools that the Stanford study is based on.

“LexisNexis has not been contacted by Stanford’s Daniel Ho [head researcher on the study], and our own data analysis suggests a much lower rate of hallucination. LexisNexis has extensive programs and system measures in place to improve the accuracy of responses over time, including the validation of citing authority references to mitigate hallucination risks in our product.

“Lexis+ AI delivers hallucination-free linked legal citations. The linked statement means that the reference can be reviewed by a user via a hyperlink. In the rare instance that a citation appears without a link, it is an indication that we cannot validate the citation against our trusted data set. This is clearly noted within the product for user awareness and customers can easily provide

feedback to our development teams to support continuous product improvement.

“LexisNexis focuses on AI answer quality through an enhanced LexisNexis proprietary Retrieval Augmented Generation 2.0 (RAG 2.0) platform. Lexis+ AI responses are grounded in an extensive repository of current, exclusive legal content which ensures the highest-quality answer with the most up-to-date validated citation references.

“The solution is continually improving with hundreds of thousands of rated answer samples by LexisNexis legal subject matter experts used for model tuning. LexisNexis employs over 2,000 technologists, data scientists, and J.D. subject matter experts to develop, test, and validate its solutions and deliver comprehensive, authoritative information.

“LexisNexis agrees with several of the generative AI legal challenges described by Professor Ho. As part of our customer feedback-driven development focus, continuous RAG development is designed to improve answer quality. The Lexis+ AI RAG 2.0 platform was released in late April and the service improvements address many of the issues noted. RAG technology is improving at an astonishing rate and users will see week over week improvements in the coming months.”