LAW.COM

Expert Opinion | **Artificial Intelligence**

# The Confidence Dashboard: Why AI Humility Beats AI Accuracy

There is a paradox at the heart of AI trustworthiness: The path to being trusted is admitting when you are uncertain. Uniformly confident AI that never acknowledges doubt is, counterintuitively, less trustworthy than the humble AI that flags its own weak spots. We would never tolerate an associate who delivered every statement with identical certainty. We should not tolerate it from our AI tools, either.

March 13, 2026 at 11:41 AM By **Michael William Ott**

## What You Need to Know

- The consequences of failing to verify AI output are no longer hypothetical.
- The path to trustworthy AI is not eliminating uncertainty but surfacing it.
- A confidence dashboard would enable verification triage is not a shortcut around professional responsibility; it is a tool that makes fulfilling that responsibility more efficient.

Imagine a junior associate who delivers every memo with identical confidence.

"The controlling case is *Smith v. Jones*," she says, whether she has spent three hours confirming the citation or is half-remembering something from law school. "The statute of limitations is two years," she announces with the same unwavering certainty, regardless of whether she checked the code or is simply guessing.

You would not tolerate this associate for long. Her uniform confidence makes her less useful, not more, because you cannot triage your review. Every assertion demands the same level of scrutiny.

Yet this is precisely how every major large language model operates today. And the artificial intelligence industry's proposed solution — spending billions to marginally reduce hallucinations — fundamentally misunderstands what lawyers actually need.

## The Verification Mandate: What Formal Opinion 512 Requires

In July 2024, the ABA Standing Committee on Ethics and Professional Responsibility issued Formal Opinion 512, the first comprehensive guidance on lawyers' use of generative artificial intelligence. The opinion's central message is unambiguous: Lawyers who use AI remain fully responsible for the accuracy of their work product.

The opinion grounds this responsibility in Model Rule 1.1's duty of competence, which requires lawyers to provide "the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation." Comment 8 to that rule, amended in 2012 and now adopted by 40 states, the District of Columbia and Puerto Rico, specifies that competence includes understanding "the benefits and risks associated with relevant technology."

Formal Opinion 512 applies these principles directly to generative AI. "Because GAI tools are subject to mistakes," the opinion states, "lawyers' uncritical reliance on content created by a GAI tool can result in inaccurate legal advice to clients or misleading representations to courts and third parties." The conclusion follows inexorably: "a lawyer's reliance on, or submission of, a GAI tool's output — without an appropriate degree of independent verification or review of its output — could violate the duty to provide competent representation as required by Model Rule 1.1."

The opinion further justifies the obligation to independently verify GAI output by citing other model rules. Rule 3.1 prohibits asserting frivolous claims; Rule 3.3 requires candor toward the tribunal; Rule 8.4(c) bars conduct involving "dishonesty, fraud, deceit, or misrepresentation." As the opinion notes, "[e]ven an unintentional misstatement to a court can involve a misrepresentation under Rule 8.4(c)." The message could not be clearer: Verify everything.

## The Growing Toll of Verification Failures

The consequences of failing to verify AI output are no longer hypothetical. Since the watershed *Mata v. Avianca* decision in June 2023 — where a New York federal judge sanctioned attorneys for submitting a ChatGPT-generated brief filled with fabricated cases — courts have documented over 846 instances of AI hallucinations in legal filings. The pace is accelerating: Researchers tracking these cases report that what was "two cases per week" in early 2025 has become "two or three cases per day."

The sanctions have grown progressively more severe. Monetary fines, once considered sufficient, are proving inadequate as a deterrent. In a July 2025 case from the Northern District of Alabama, the court declared that it was moving beyond fines to

disqualification — removing offending attorneys from the case entirely. An Arizona magistrate judge revoked an attorney's pro hac vice status, struck her brief, ordered apology letters to judges whose names were falsely attributed to nonexistent cases, and referred her to the state bar.

These failures span the profession. A Colorado federal judge fined attorneys representing Mike Lindell $3,000 each for a filing with "more than two dozen mistakes — including hallucinated cases." A Utah attorney was ordered to pay attorney fees, refund client fees and donate $1,000 to a legal nonprofit. A Denver attorney accepted a 90-day suspension after admitting he hadn't checked ChatGPT's work.

What makes these cases particularly troubling is their diversity. The attorneys involved range from solo practitioners to partners at well-regarded firms. Some had adopted internal AI policies; others had not. But all shared a common failure: insufficient verification of AI-generated content.

## The Overconfidence Problem

Why do these failures keep occurring? Part of the answer lies in how AI systems present their outputs.

Large language models generate text by predicting the most probable next token (roughly, the next word) based on patterns learned from training data. Internally, the model has varying degrees of confidence in these predictions. Some outputs reflect high-probability predictions drawn from consistent, extensive training data; others are low-probability guesses stitched together to sound plausible.

But users never see this internal uncertainty. A hallucinated case citation is delivered with the same authoritative tone as black-letter law. The model's confident voice masks its actual confidence level — creating a dangerous mismatch between presentation and reliability.

This is the equivalent of the uniformly confident associate described above. When everything sounds equally certain, verification becomes binary: Check everything or trust nothing. There is no middle ground, no ability to allocate review effort intelligently and efficiently.

For lawyers bound by ethical duties of competence and candor, this creates an impossible workflow. Formal Opinion 512 acknowledges that "the appropriate amount of independent verification or review required to satisfy Rule 1.1 will necessarily depend on the GAI tool and the specific task." But current AI tools provide no help in making that determination.

## A Better Investment: The Confidence Dashboard

Here is the counterintuitive insight: The path to trustworthy AI is not eliminating uncertainty but surfacing it.

Consider the user interface of any major AI chatbot. Below each output, you will find buttons: copy, thumbs up, thumbs down, regenerate. What if there were one more button — a "confidence scorecard" or "confidence dashboard" that extracted the major assertions from the output and displayed a confidence score for each?

The technical barriers are minimal. The probability distributions are already generated as part of the inference process; they simply are not exposed to users. Surfacing them requires user interface development and some additional processing but nothing approaching the billions required to retrain models for marginal accuracy gains.

The user benefit would be transformative. A confidence dashboard would enable verification triage — exactly what Formal Opinion 512 contemplates when it acknowledges that the "appropriate degree of independent verification" is context-dependent. High-confidence assertions (say, above 90%) might warrant a quick spot-check. Medium-confidence items (60-90%) would receive normal verification. Low-confidence claims (below 60%) would trigger deep research or immediate rejection.

The lawyer still verifies everything — as Model Rule 1.1 requires — but can now allocate time intelligently rather than treating every assertion as equally suspect. This is not a shortcut around professional responsibility; it is a tool that makes fulfilling that responsibility more efficient.

## The Economics: Verification Infrastructure Versus Accuracy Obsession

The AI industry is currently spending billions chasing marginal improvements in accuracy. The goal seems to be reducing hallucination rates from, say, 3% to 1%. But consider what this means for a lawyer who must comply with Formal Opinion 512.

Whether the AI hallucinates 3% of the time or 1% of the time, the lawyer must still verify every material assertion before relying on it in client work or court filings. The 2% improvement in accuracy does not change the verification workflow — it just means slightly fewer corrections after the review process that was going to happen anyway.

Now compare two investment strategies: strategy A, spend billions on compute and training to reduce hallucination rates from 3% to 1%; strategy B, spend millions on verification infrastructure — confidence dashboards, citation tools, reasoning traces — to make human review faster and more targeted.

For a lawyer bound by the duties articulated in Formal Opinion 512, strategy B delivers far more value. The accuracy improvement changes nothing about the verification requirement; the verification infrastructure transforms how efficiently that requirement can be met.

An AI tool with a 3% hallucination rate and excellent verification infrastructure — one that flags its own uncertainty, provides confidence scores and highlights assertions that warrant extra scrutiny — may be more valuable than a tool with a 1% hallucination rate that presents everything with equal, unjustified confidence.

## Implications for Practice

- **For lawyers evaluating AI tools:** The relevant question is not just "how accurate is it?" but "how easy is it to verify?"[A1] When assessing AI tools for legal work, consider whether the tool surfaces its uncertainty, provides citation support or offers other verification infrastructure. Raw accuracy metrics tell an incomplete story.
- **For AI providers:** The legal market represents a substantial user base with mandatory verification requirements that will not change, regardless of accuracy improvements. Confidence dashboards, citation verification tools and reasoning traces may deliver a better return on investment than marginal accuracy gains. Building "humble AI" that acknowledges its limitations may be more valuable than building confident AI that masks them.
- **For bar associations and regulators:** Future guidance on AI use in legal practice should consider not just accuracy metrics but verification infrastructure. When evaluating whether lawyers are meeting their competence obligations, consider whether the tools they use facilitate or impede the verification that Formal Opinion 512 requires.
- **For law firms developing AI policies:** Training should emphasize not just the risks of AI hallucinations but the importance of selecting tools that support verification workflows. Policies that focus solely on prohibiting AI use miss the opportunity to leverage tools that make verification more efficient.

## The Humility Paradox

There is a paradox at the heart of AI trustworthiness: The path to being trusted is admitting when you are uncertain. Uniformly confident AI that never acknowledges doubt is, counterintuitively, less trustworthy than the humble AI that flags its own weak spots.

We would never tolerate an associate who delivered every statement with identical certainty. We should not tolerate it from our AI tools, either.

The technology to solve this problem already exists — it is simply not exposed to users. The economics favor verification infrastructure over accuracy obsession. And the professional obligations articulated in Formal Opinion 512 are not going away; if anything, they will only intensify as AI becomes more integrated into legal practice.

AI providers that recognize this will build confidence dashboards and verification tools. The lawyers who demand these features will get better tools. And the profession that embraces AI humility will, paradoxically, place more justified trust in its AI-assisted work product.

*****

**Michael William Ott** is a bankruptcy partner at Ice Miller LLP in Chicago, where he serves on the firm's Technology Adoption Committee and focuses on AI implementation strategy in legal practice.

***This article originally appeared in Accounting and Financial Planning for Law Firms, a Law Journal Newsletters publication covering all financial aspects of managing law firms, including: building a law firm budget; rates and rate arrangements with clients; coordinating benefits for law firm partners; and the newest strategies to grow your firm and your career.***