CLIENT ALERTS & INSIGHTS

When ChatGPT Lies: What the First Wave of AI Defamation Cases Means for Plaintiffs

JUNE 24, 2026

Authors: David D. Pope, Maura T. Levine-Patton

Featured Practices: Litigation, Defamation Litigation, Crisis Management & Strategic Response

Key Takeaways

- AI tools are already generating false, reputation-damaging statements about real people-but early defamation lawsuits against companies like OpenAI, Google and Meta are struggling because courts are applying traditional defamation rules and often finding no liability (especially where disclaimers exist or harm isn't proven).
- There is a clear disconnect between real-world reputational harm and current legal remedies, meaning individuals and businesses may have limited recourse when AI "hallucinates," while companies face ongoing but still uncertain liability exposure as courts hesitate to expand doctrine.
- Plaintiffs will need to build stronger, evidence-driven cases-focusing on proof of real-world harm, third-party reliance, repeat errors and platform notice-while also considering alternative legal theories beyond traditional defamation as this area of law evolves.

Artificial intelligence platforms are generating false-and sometimes seriously damaging-statements about real people. Recent lawsuits against Google, Meta and OpenAI confirm that AI "hallucinations" are no longer hypothetical; they are landing in court. But early litigation is struggling to gain traction. Traditional defamation doctrine does not map neatly onto AI-generated speech, and courts remain reluctant to expand liability. For plaintiffs' lawyers, understanding where these claims stand today is essential to shaping the cases of tomorrow.

What Are Courts Grappling With?

Across recent cases, the core fact pattern is remarkably consistent: a chatbot generates false statements about a real person-accusations of criminal conduct, professional misconduct or extremist affiliations-and presents them in an authoritative format such as a biography, summary or direct answer to a user query. The output either lacks any traceable source or cites fabricated ones.

Plaintiffs argue this is not merely "wrong information" but classic defamation: a false statement of fact communicated to a third party that would damage the subject's reputation if believed. In *Walters v. OpenAI*

, radio host Mark Walters sued after ChatGPT told a user that Walters had embezzled funds from a Second Amendment advocacy organization. *Walters v. OpenAI, LLC*, No. 23-A-04860-2 (Ga. Super. Ct., May 19, 2025); *see also Starbuck v. Google, LLC*, No. N25C-10-211 (Del. Super. Ct., Oct. 22, 2025) (plaintiff sued Google after its chatbot generated false accusations of sexual assault and criminal records).

Early Judicial Signal: Courts Are Skeptical-For Now

The most developed decision to date, *Walters v. OpenAI, LLC*, signals that courts may resist holding companies liable for chatbot hallucinations on familiar grounds: (1) the content lacked "defamatory meaning," and (2) the defendant lacked the requisite state of mind.

On the first point, the *Walters* court held that ChatGPT's output did not carry defamatory meaning as a matter of law because users are explicitly warned that outputs may be inaccurate. In the court's view, OpenAI's disclaimer effectively refuted any claim that a reasonable reader would interpret the chatbot's statements as assertions of fact.

On the second point, the court found that OpenAI did not act with even ordinary negligence-the baseline standard of whether a reasonable publisher would have acted differently. OpenAI's expert testified that the company "leads the AI industry in attempting to reduce and avoid mistaken output like the challenged output here." *Walters v. OpenAI, LLC*, 2025 WL 2979163, at *5 (Ga. Super. Ct., May 19, 2025). The court was unwilling to penalize a company for the very problem it was actively trying to solve.

The practical takeaway: even when the defendant is an AI company, courts are still applying traditional defamation elements-falsity and scienter-as threshold gatekeepers. At this stage of the technology, imposing liability for hallucinations could amount to strict liability, and courts appear unwilling to go there. Unlike a human professional who should know better, a chatbot specifically disclaims the accuracy of its output. Until the technology matures, courts will likely continue to find that users knew or should have known the results could be wrong, and therefore should not have taken the output at face value.

Publication and Republication

Significant questions remain unanswered, particularly around publication. Defamation, of course, hinges on the communication of a false statement to a third party. Both original publishers and republishers (those who know a statement is false but spread it further) face liability. AI complicates that framework. If a chatbot generates defamatory content only in response to a specific user prompt, who "published" it: the platform, the user who typed the query or no one?

A related question is whether a plaintiff's claim weakens when the person who received the false output does not pass it along to anyone else. In *Walters*, journalist Frederick Riehl was researching a lawsuit when he prompted ChatGPT to summarize the case, triggering the allegedly defamatory output about plaintiff Mark Walters. *Walters,* 2025 WL 2979163, at *2. Riehl never republished the information, and Walters himself testified he suffered no actual damage from the false statement. *Id.* Those facts mattered. Had Riehl republished ChatGPT's output without confirming it, could he have faced liability as a republisher?

The upshot: evidence of third-party exposure, reliance and actual damage will be critical for plaintiffs looking to prove dissemination and reputational harm. With precedent still thin and the landscape shifting**,** plaintiff-side practitioners should consider several concrete strategies:

1. Build the "Real-World Impact" Record Early

Courts are sensitive to whether false statements were:

- Seen
- Believed
- Acted upon

The goal is to document reputational harm beyond the chatbot output itself: lost business, withdrawn invitations, damaged relationships.

2. Focus on Repeat Conduct and Notice

A single, isolated hallucination is a hard case. Stronger claims will involve:

- Repeated false outputs
- Prior notice to the platform
- Failure to correct or remediate

3. Control the Narrative Around Reasonableness

Defendants will argue that "no reasonable person would rely on AI." Plaintiffs should be prepared to counter with evidence showing:

- How consumers actually use these tools
- The increasing reliance on AI as an information source
- The authoritative presentation of outputs

4. Consider Claims Beyond Traditional Defamation

Where traditional defamation elements are difficult to establish, plaintiffs may explore alternative theories:

- False light
- Negligence-based theories
- Product liability or consumer protection frameworks (fact-dependent)

The Bottom Line

AI defamation claims are no longer theoretical-but liability has not yet caught up with the harm. Early decisions suggest that courts are:

- Hesitant to treat AI output as conventional "publication"
- Skeptical of reliance where disclaimers exist
- Reluctant to impose broad liability on developers absent clear fault

But litigation is accelerating. The underlying facts-false, reputationally damaging statements generated by widely used platforms-are compelling, and courts will not be able to avoid these questions indefinitely. For plaintiffs, the path forward is not foreclosed, but it demands a deliberate, evidence-driven approach built on real-world impact.

Benesch's Crisis Management & Defamation team continues to actively monitor the implications of AI as it relates to defamation litigation across the country. Benesch client alerts and legal publications are available for you to receive by signing up HERE.