# Anthropic’s Claude AI cited fake legal source in copyright case, sparking trust concerns



Anthropic, a leading AI firm, recently faced significant scrutiny after its AI model, Claude, generated a fabricated legal citation that was subsequently used in a copyright lawsuit. This incident has raised serious questions about the reliability of AI in high-stakes professional environments, particularly in the legal sector. Following the revelation that the erroneous citation stemmed from Claude’s output, Anthropic formally apologised, labelling the situation an “embarrassing and unintentional mistake.” The misstep came to light during a legal challenge initiated by music publishers who claimed that Claude was trained unlawfully using their copyrighted song lyrics.

The incident underscores broader concerns regarding the accuracy of AI-generated information, particularly in situations where precise details are paramount. Details emerged when Ivana Dukanovic, an associate at the law firm Latham & Watkins, tasked Claude with generating a properly formatted legal citation. Unfortunately, while the AI provided a valid link to an academic source, it included an incorrect article title and inaccurately attributed authorship. This was all overlooked during the law firm's manual review process, which serves as a stark reminder of the crucial role human oversight must play in validating AI outputs.

U.S. Magistrate Judge Susan van Keulen, who presided over the case, highlighted the gravity of such a misrepresentation, asserting that the generation and use of a fabricated citation constitutes a “very serious and grave issue.” During the proceedings, the plaintiffs' attorney, Matt Oppenheim, revealed that he had confirmed with one of the cited authors that no such article existed, suggesting that the erroneous citation potentially enabled the usage of Claude to bolster legal arguments within the case.

In the wake of these mistakes, Anthropic’s attorney attempted to downplay the error, arguing it was merely a mis-citation rather than a full fabrication. Judge van Keulen, however, established a clear distinction, noting the significant implications of AI-generated “hallucinations”, which often result in the wrongful dissemination of false information. The music publishers have also called for sanctions against the attorneys from Latham & Watkins due to this oversight.

This incident is not an isolated case; rather, it reflects a trend seen across the legal profession where AI tools have led to significant inaccuracies in documentation. Recent discussions have revealed that legal practitioners who rely on AI must navigate a complex landscape marked by the potential for misinformation. Similar criticisms have been voiced against law firms utilizing AI tools, with a notable report from The Verge highlighting cases in which judges have condemned “bogus AI-generated research.”

The ramifications of these reliability issues extend beyond courtroom implications. Even as AI firms, including Anthropic, implement new internal procedures aimed at reducing citation errors, the very nature of AI 'hallucinations' persists as a challenge. Meta’s AI, for instance, faced backlash for misrepresenting a widely reported news event, prompting officials to acknowledge the prevalence of hallucinations across generative AI systems.

Some industry experts underscore the responsibility of legal professionals to verify AI-generated outputs thoroughly. As reflected in a recent study presented at the CHI 2025 conference, there is a worrying trend where users may unknowingly prefer AI-generated legal advice, ignoring its inherent unreliability. This raises substantial concerns about potential misuse of flawed information, driving the call for robust oversight and verification mechanisms.

Furthermore, as the legal community grapples with these challenges, AI research initiatives are underway to develop solutions. The SAFE (Search-Augmented Factuality Evaluator) system, introduced by researchers from Google DeepMind and Stanford University, aims to enhance the accuracy of AI responses by cross-checking generated facts against reliable sources.

The current landscape suggests that AI technology, while innovative, requires stringent protocols to ensure its safe integration into high-stakes sectors like law. As both businesses and regulatory bodies consider ways to address these complications, the conversation continues about establishing transparency mandates and usage guidelines, similar to those proposed in the EU AI Act. Only through a commitment to responsible AI use and rigorous human verification can the sector hope to maintain public trust and accountability as generative AI tools evolve.

### Reference Map

1. Paragraphs 1, 2, 3, 4
2. Paragraph 3
3. Paragraphs 5, 6
4. Paragraph 8
5. Paragraph 8
6. Paragraph 9

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://winbuzzer.com/2025/05/15/anthropics-lawyers-apologize-after-its-claude-ai-hallucinates-legal-citation-in-copyright-lawsuit-xcxwbn/> - Please view link - unable to able to access data
2. <https://www.reuters.com/legal/legalindustry/anthropics-lawyers-take-blame-ai-hallucination-music-publishers-lawsuit-2025-05-15/> - In a recent copyright lawsuit filed by music publishers against AI firm Anthropic, the company's legal team admitted responsibility for an erroneous citation in an expert report. The mistake originated from a fabricated citation generated by Anthropic's AI chatbot, Claude, which provided an incorrect article title and authors despite linking to a valid academic source. This AI 'hallucination' has raised concerns with U.S. Magistrate Judge Susan van Keulen, who emphasized the serious implications of such errors in legal proceedings. The incident underscores the challenges of integrating AI tools into professional environments and the necessity for meticulous human oversight. ([reuters.com](https://www.reuters.com/legal/legalindustry/anthropics-lawyers-take-blame-ai-hallucination-music-publishers-lawsuit-2025-05-15/?utm_source=openai))
3. <https://www.reuters.com/legal/litigation/anthropic-expert-accused-using-ai-fabricated-source-copyright-case-2025-05-13/> - A federal judge in San Jose, California, has ordered Anthropic to respond to accusations that it used a fabricated source generated by AI in a copyright lawsuit filed by music publishers, including Universal Music Group, Concord, and ABKCO. The lawsuit centers on the alleged misuse of copyrighted song lyrics to train Anthropic’s AI chatbot, Claude. During a hearing, the plaintiffs’ attorney, Matt Oppenheim, claimed that an Anthropic data scientist cited a nonexistent academic article to support arguments about how frequently Claude reproduces copyrighted lyrics. The article was allegedly generated by Anthropic’s AI and falsely attributed to a respected journal. Judge Susan van Keulen labeled the incident a serious issue and demanded a prompt response from Anthropic, though she denied an immediate deposition of the expert, Olivia Chen. Anthropic acknowledged a citation error but suggested it might relate to a different, legitimate article. The case underscores growing concerns over AI-generated misinformation in legal documents, as other attorneys have recently faced sanctions for similar hallucinations. ([reuters.com](https://www.reuters.com/legal/litigation/anthropic-expert-accused-using-ai-fabricated-source-copyright-case-2025-05-13/?utm_source=openai))
4. <https://www.reuters.com/legal/anthropic-wins-early-round-music-publishers-ai-copyright-case-2025-03-26/> - Anthropic, an artificial intelligence company, achieved a legal victory when a federal judge in California dismissed a preliminary request to prevent the use of song lyrics owned by Universal Music Group and other music publishers for training its chatbot, Claude. The judge noted that the publishers failed to demonstrate 'irreparable harm' and that their request was overly broad. The publishers had filed a lawsuit alleging that Anthropic used their lyrics without permission to train Claude, marking a significant case in the ongoing debate over AI's use of copyrighted material. ([reuters.com](https://www.reuters.com/legal/anthropic-wins-early-round-music-publishers-ai-copyright-case-2025-03-26/?utm_source=openai))
5. <https://www.reuters.com/legal/litigation/anthropic-reaches-deal-ai-guardrails-lawsuit-over-music-lyrics-2025-01-03/> - Anthropic, an artificial intelligence company, has reached an agreement with Universal Music and other music publishers to implement 'guardrails' to prevent its chatbot, Claude, from generating copyrighted song lyrics. This deal, approved by the U.S. District Judge Eumi Lee, partially resolves a lawsuit accusing Anthropic of using song lyrics from artists like Beyoncé and the Rolling Stones without permission to train Claude. Although Anthropic has denied the allegations and claims it already had measures in place, it agreed to maintain and extend these guardrails to future models. The publishers' broader request for a preliminary injunction is still under consideration. This case is part of a wider trend of lawsuits against AI companies for allegedly using copyrighted material without consent. ([reuters.com](https://www.reuters.com/legal/litigation/anthropic-reaches-deal-ai-guardrails-lawsuit-over-music-lyrics-2025-01-03/?utm_source=openai))
6. <https://www.reuters.com/legal/litigation/anthropic-asks-court-dismiss-music-publishers-ai-claims-2024-08-16/> - Artificial intelligence company Anthropic has requested that a California federal court dismiss some copyright claims made by music publishers over the alleged misuse of song lyrics to train its AI chatbot, Claude. The publishers, including Universal Music Group, ABKCO, and Concord Music Group, accuse Anthropic of infringing on their copyrights by using their lyrics without authorization. Anthropic denies inducing users to infringe copyrights or committing other violations, arguing that the publishers' secondary claims are implausible. However, the company refrains from addressing the core issue of whether training AI with copyrighted lyrics constitutes fair use. The ongoing lawsuit, filed last October, is notable for being the first to focus on song lyrics and involves lyrics from at least 500 songs by artists such as Beyoncé and the Rolling Stones. Additionally, the publishers seek a court order to stop Anthropic's use of their lyrics pending resolution. ([reuters.com](https://www.reuters.com/legal/litigation/anthropic-asks-court-dismiss-music-publishers-ai-claims-2024-08-16/?utm_source=openai))