Knowledge Management

ALERT | 20 June 2025

CDH

In this issue

SOUTH AFRICA

Formatting with AI may be riskier than you realise

Formatting with AI may be riskier than you realise

In *Concord Music Group, Inc. v Anthropic PBC*, 5:24-cv-03811, (N.D. Cal.) lawyers thought they did everything right. They researched legitimate sources, found real academic papers, and only asked an artificial intelligence (AI) tool to help format their citations. Yet a federal judge still struck their evidence for containing fabricated references. Welcome to AI's latest trick: corrupting good research.

The unexpected error

The dispute arose from a request for data sampling in copyright litigation against Anthropic, the maker of Claude. Both parties submitted expert declarations with competing proposals for reviewing millions of prompt-output pairs. During the hearing, opposing counsel challenged Anthropic's expert, asking the court to strike her declaration because citations appeared to reference articles that didn't exist with authors who had never worked together.

The investigation revealed something more nuanced than typical AI fabrication. Anthropic's counsel claimed they had conducted proper human-led research, locating legitimate academic sources through Google searches. They then used Claude to "*properly format*" the citations for submission. In the process, Claude generated partially fictitious citations.

The declaration contained reference to a paper titled "*Binomial Confidence Intervals for Rare Events*". Although Claude used the correct publication year, and a link to the provided source, the returned citation included an inaccurate title and incorrect authors. Claude also introduced two other formatting errors: changing "*Computing Necessary Sample Size*" to "*Sample Size Estimation*" in another citation and adding the word "*Lower*" to "*Windward Environmental LLC*" in a third reference.

Despite conducting a manual citation check, the legal team failed to catch these errors before submission.

Not your typical AI hallucination

This case represents something distinct from the fabricated cases we've seen in courts. The lawyers conducted proper research and found legitimate sources. The error occurred when AI was asked to perform what seemed like a simple formatting task and was compounded when the manual review process proved lacking.

Judge van Keulen called it a "*plain and simple AI hallucination*", though this characterisation may not capture the full picture. Technically, AI language models don't actually "*format*" citations: they generate new text based on patterns in training data. When asked to format a citation, Claude likely generated what it predicted a properly formatted citation should look like rather than simply reformatting the provided information. This generative process can introduce errors even when working with accurate source material.

Formatting with AI may be riskier than you realise

CONTINUED

Ultimately, the court recognised this wasn't a case where "*attorneys and experts abdicated their independent judgment and critical thinking skills in favour of ready-made, AI-generated answers*". However, the verification failure remained "*a serious concern*", particularly given how obvious the errors were once identified.

The court's response

The judge's ruling was measured but firm. She struck the affected portion and explicitly stated that the incident "*undermines the overall credibility of* [Anthropic's expert's] *written declaration*". While stopping short of sanctions, the court expressed bewilderment at how manual verification could miss such fundamental mistakes.

The ruling illustrates that courts will not tolerate verification failures, regardless of whether AI tools were used with good intentions or for seemingly simple tasks.

Rethinking verification in the age of AI

The most significant takeaway is that verification protocols cannot be superficial. It's insufficient to simply check that sources exist. Every AI-generated output requires human verification. In this case, the team verified their research but failed to scrutinise the AI's formatting output.

Legal practice has developed sophisticated verification methods over centuries, carefully calibrated to catch predictable human errors in traditional drafting processes. Generative AI introduces new categories of errors that our established review techniques weren't designed to identify, and the incident highlights a broader cautionary principle: users should approach AI-generated output with heightened vigilance, particularly when referencing sources. The ostensible efficiency gains from AI tools can prove illusory when errors require extensive correction and professional embarrassment follows. Smart deployment of AI tools requires understanding their limitations and recognising when traditional methods remain more reliable.

The legal profession is still developing best practices that properly balance AI's efficiency advantages against the heightened verification demands it creates. As a starting point, lawyers who use AI should recognise its generative nature. AI models generate text by predicting the most likely next words or sequences based on patterns learned from vast amounts of training data. While AI often provides appropriate and accurate responses, it does not "*copy and paste*" or systematically reorganise input data, but rather generates new text that approximates a desired style. And because the output is generated probabilistically, errors can occur.

Formatting with AI may be riskier than you realise

CONTINUED

The next step is recognising that AI excels at making fabricated content appear legitimate. The fabricated citations in this case would have looked entirely professional to casual observers. Lawyers must treat AI-generated content with healthy scepticism, particularly for court submissions where accuracy is paramount. In this case, a simple Google search for proper citation formatting would have equipped the legal team with the knowledge to handle the task manually, avoiding the risk entirely.

Finally, law firms should ensure that there are clear protocols in place for verification. While court submissions and marketing publications may not each require the same level of scrutiny, implementing systematic verification processes will safeguard firms' reputations and, more importantly, preserve the interests of their clients.

Takeaways for South African lawyers

In *Concord*, the court's measured but firm response demonstrates that verification failures carry real consequences, regardless of intent. While this US case may not have precedential value here, it proves highly relevant given South Africans' own experiences with AI hallucinations in legal proceedings (for example, in the recent case of *Mavundla v MEC Department of Co-Operative Government and Traditional Affairs and Others* [2025] ZAKZPHC 2).

Crucially, our courts haven't yet established clear guidelines on AI use in legal practice. This absence of specific directives means lawyers must apply the strictest verification standards themselves, proactively rather than reactively.

The case underscores two critical professional obligations in the AI era: maintaining detailed records of how AI tools are used in legal work, and immediately disclosing mistakes when they occur. Transparency about AI involvement and prompt acknowledgment of errors can help preserve professional relationships and credibility, even when verification processes fail.

The takeaway isn't to abandon AI tools but to use them more intelligently. Establish comprehensive verification protocols, understand AI's generative nature, and maintain human oversight proportionate to the stakes involved. In an era where courts are increasingly alert to AI-related errors, robust verification is essential for maintaining professional credibility and serving clients effectively.

Retha Beerman, Safee-Naaz Siddiqi and Shanley Webb

JOHANNESBURG

1 Protea Place, Sandton, Johannesburg, 2196. Private Bag X40, Benmore, 2010, South Africa.
Dx 154 Randburg and Dx 42 Johannesburg.
T +27 (0)11 562 1000   F +27 (0)11 562 1111   E jhb@cdhlegal.com

CAPE TOWN

11 Buitengracht Street, Cape Town, 8001. PO Box 695, Cape Town, 8000, South Africa. Dx 5 Cape Town.
T +27 (0)21 481 6300   F +27 (0)21 481 6388   E ctn@cdhlegal.com

NAIROBI

Merchant Square, 3rd floor, Block D, Riverside Drive, Nairobi, Kenya. P.O. Box 22602-00505, Nairobi, Kenya.
T +254 731 086 649 | +254 204 409 918 | +254 710 560 114
E cdhkenya@cdhlegal.com

NAMIBIA

1st Floor Maerua Office Tower, Cnr Robert Mugabe Avenue and Jan Jonker Street, Windhoek 10005, Namibia
PO Box 97115, Maerua Mall, Windhoek, Namibia, 10020
T +264 833 730 100 E cdhnamibia@cdhlegal.com

STELLENBOSCH

14 Louw Street, Stellenbosch Central, Stellenbosch, 7600.
T +27 (0)21 481 6400 E cdhstellenbosch@cdhlegal.com