

Dispute Resolution and Knowledge Management

ALERT | 10 June 2025



In this issue

SOUTH AFRICA

When AI lies: The legal minefield of artificial intelligence and defamation



For more insight into our
expertise and services

When AI lies: The legal minefield of artificial intelligence and defamation

In a world where artificial intelligence can write poetry, diagnose diseases and draft legal documents, what happens when it tells lies about a person? This question moved from theory to reality when OpenAI, through its AI chatbot ChatGPT, found itself at the centre of a legal battle that should serve as a warning to those who blindly rely on artificial intelligence as a source of information.

The case began innocuously enough: a journalist asked ChatGPT to summarise a legal complaint against the Attorney General of the state of Washington, laid by the Second Amendment Foundation, for which Walters was a spokesperson (the Ferguson complaint). For those who may still be uncertain of how ChatGPT works, the court provided a helpful summary:

"ChatGPT is a tool that allows users to access a 'large language model' or 'LLM' that is 'trained on vast amounts of data' to generate new text in response to a user's prompt 'by predicting what words will come next', and 'due to their generative nature, all of the major LLMs that are currently available to the public' are capable of 'generat[ing] information contradicting the source material,' sometimes referred to as 'hallucinations'."

While ChatGPT initially summarised extracts of the complaint correctly, its subsequent summary of information available in an internet link (URL) to the complaint contained a serious allegation – ChatGPT provided a new, incorrect summary of the Ferguson complaint, claiming that Walters had embezzled funds from a gun rights organisation, a statement with no basis in reality. It was a classic AI hallucination, but one which could have seriously impacted Walters' reputation if believed.

Fortunately, the journalist researched the Ferguson complaint further and quickly established that the ChatGPT output was not true and no further publication of the fabricated allegations ensued.

The Georgia precedent

In *Walters v OpenAI*, L.L.C, No. 23-A-04860-2 (Ga. Super. Ct. 2025) Walters sued OpenAI for defamation, arguing that ChatGPT's false statement about embezzlement constituted publication of defamatory material. The Superior Court of Gwinnett County, Georgia, had to grapple with a novel question: can an AI developer be held liable when its technology generates defamatory content?

The case presented several unprecedented legal issues. Could AI-generated hallucinations be attributed to OpenAI as intentional or negligent publication? Do traditional defamation standards applicable in the US – requiring intent, negligence or actual malice – apply to algorithmic outputs that are inherently unpredictable? And, crucially, in an age where AI companies plaster their products with warnings about potential inaccuracies, can users reasonably rely on AI-generated statements as fact?

When AI lies: The legal minefield of artificial intelligence and defamation

CONTINUED

The Georgia court ultimately granted summary judgment in favour of OpenAI. The decision hinged on several key findings, namely that (i) ChatGPT's output was not defamatory as a matter of law; (ii) Walters could not prove negligence or actual malice on the part of OpenAI; and (iii) he suffered no quantifiable damages. Perhaps most significantly, the court determined that no reasonable reader of the ChatGPT output, who had experience with the AI tool and who would have received repeated disclaimers which warned that mistaken output was a real possibility, would interpret ChatGPT's output as stating "actual facts".

The court emphasised OpenAI's proactive steps to reduce errors and provide robust warnings about AI hallucinations, which were included in its terms of use. These safeguards, combined with the probabilistic nature of AI technology, created a legal shield that protected the company from liability.

The South African perspective

How might this case have unfolded in South Africa, where the principles of defamation differ from those applicable in the US?

South African defamation law requires the wrongful and intentional publication of a defamatory statement pertaining to a person. To succeed with a defamation claim, a plaintiff has to prove the following:

- a defamatory statement;
- its publication;
- that the statement was made of and concerning the plaintiff; and
- that the statement has a defamatory meaning to a reasonable reader.

Once a plaintiff has proved the above elements, it is presumed that the publication is unlawful and that the defendant acted with the intention to defame the plaintiff. The onus then shifts to the defendant to rebut these presumptions of unlawfulness.

When considering the defences that can be raised by defendants, i.e. recognised grounds for lawful justification, a distinction can be drawn between media defendants and other defendants, in that the media can also rely on the so-called reasonable publication defence, established in the landmark case of *National Media Ltd v Bogoshi* 1996 (3) SA 78. The reasonable publication defence entails that a media defendant will not be held liable for the publication of false and defamatory statements, if in publishing such statements, it did not act negligently and the publication of the defamatory material was reasonable considering all the circumstances.

With this in mind, had the journalist published ChatGPT's defamatory fabrication about Walters, he would likely have been held liable under South African law for his failure to take reasonable steps to verify the accuracy of the information. Not only was the defamatory content not true, but the journalist would not have acted reasonably in publishing the content. The court would have considered the circumstances, including the fact that the journalist had first-hand experience of hallucinations by AI tools, as well as the numerous disclaimers that would have come to the journalist's attention, and determined that the journalist did not act reasonably in publishing the defamatory content.

When AI lies: The legal minefield of artificial intelligence and defamation

CONTINUED



The more intriguing question is whether defamatory AI hallucinations can and should be attributed to the AI developer for purposes of defamation.

The probabilistic nature of AI algorithms should make it rather easy for an AI developer to prove that it had no intention to defame a person. The absence of human intervention in AI output complicates any defamation claim brought by someone based on AI hallucinations. What a court might have to consider is whether the AI developer acted reasonably in designing, deploying or providing warnings about potential hallucinations. If clear disclaimers and reasonable safeguards were implemented, liability on the side of the AI developer would be unlikely. However, if a plaintiff could demonstrate that the developer failed to implement reasonable protections or ignored foreseeable risks, liability might arise.

Broader implications

All of the above is, of course, speculative.

Were a similar case to come before a South African court, the court could decide to deal with the matter in terms of the current principles of defamation law, or it could adopt a completely new standard or take a more unexpected approach. The Walters case illuminates a fundamental challenge facing legal systems worldwide: how to adapt centuries-old defamation principles to revolutionary technology that operates in ways human publishers never could.

Defamation law principles assume human agency: someone publishes or acts negligently. AI systems operate probabilistically, generating outputs based on statistical

patterns rather than intention. This creates unprecedented questions about foreseeability and responsibility in the chain from AI development to publication. The Walters case demonstrates that robust warnings and reasonable safeguards can protect AI developers from liability, but it also highlights the growing responsibility of users to verify AI-generated information.

The legal uncertainty extends beyond defamation. If AI companies can shield themselves behind disclaimers and the inherent unpredictability of their technology, would there be sufficient incentives to improve accuracy? Conversely, if courts impose strict liability, will it stifle AI development and limit beneficial applications?

For legal systems, the case reveals the urgent need to develop coherent frameworks for AI liability. The current principles applicable to defamation claims may suffice for now, but as AI capabilities expand, the challenge for legal systems everywhere is finding the balance that protects both technological progress and human dignity in our increasingly AI-mediated world.

Moving forward, stakeholders across the AI ecosystem – from developers implementing robust safeguards to legal practitioners advising clients on emerging risks – should collaborate to establish best practices while courts and legislators work to clarify liability standards, especially as AI becomes increasingly sophisticated and ubiquitous. The challenge lies not in prescriptive solutions, but in fostering responsible innovation within an evolving legal landscape.

Anja Hofmeyr, Safee-Naaz Siddiqi, and Shanley Webb

OUR TEAM

For more information about our Dispute Resolution practice and services in South Africa, Kenya and Namibia, please contact:



Rishaban Moodley

Practice Head & Director:
Dispute Resolution
Sector Head:
Gambling & Regulatory Compliance
T +27 (0)11 562 1666
E rishaban.moodley@cdhlegal.com



Tim Fletcher

Chairperson
Director: Dispute Resolution
T +27 (0)11 562 1061
E tim.fletcher@cdhlegal.com

Imraan Abdullah

Director:
Dispute Resolution
T +27 (0)11 562 1177
E imraan.abdullah@cdhlegal.com

Timothy Baker

Director:
Dispute Resolution
T +27 (0)21 481 6308
E timothy.baker@cdhlegal.com

Eugene Bester

Director:
Dispute Resolution
T +27 (0)11 562 1173
E eugene.bester@cdhlegal.com

Neha Dhana

Director:
Dispute Resolution
T +27 (0)11 562 1267
E neha.dhana@cdhlegal.com

Denise Durand

Director:
Dispute Resolution
T +27 (0)11 562 1835
E denise.durand@cdhlegal.com

Claudette Dutilleux

Director:
Dispute Resolution
T +27 (0)11 562 1073
E claudette.dutilleux@cdhlegal.com

Jackwell Feris

Sector Head:
Industrials, Manufacturing & Trade
Director: Dispute Resolution
T +27 (0)11 562 1825
E jackwell.feris@cdhlegal.com

Nastascha Harduth

Sector Head: Corporate Debt,
Turnaround & Restructuring
Director: Dispute Resolution
T +27 (0)11 562 1453
E n.harduth@cdhlegal.com

Anja Hofmeyr

Director:
Dispute Resolution
T +27 (0)11 562 1129
E anja.hofmeyr@cdhlegal.com

Corné Lewis

Director:
Dispute Resolution
T +27 (0)11 562 1042
E corne.lewis@cdhlegal.com

Nomlayo Mabhena-Mlilo

Director:
Dispute Resolution
T +27 (0)11 562 1743
E nomlayo.mabhena@cdhlegal.com

Sentebale Makara

Director:
Dispute Resolution
T +27 (0)11 562 1181
E sentebale.makara@cdhlegal.com

Vincent Manko

Director:
Dispute Resolution
T +27 (0)11 562 1660
E vincent.manko@cdhlegal.com

Khaya Mantengu

Director:
Dispute Resolution
T +27 (0)11 562 1312
E khaya.mantengu@cdhlegal.com

Richard Marcus

Director:
Dispute Resolution
T +27 (0)21 481 6396
E richard.marcus@cdhlegal.com

Burton Meyer

Director:
Dispute Resolution
T +27 (0)11 562 1056
E burton.meyer@cdhlegal.com

Desmond Odhiambo

Partner | Kenya
T +254 731 086 649
+254 204 409 918
+254 710 560 114
E desmond.odhiambo@cdhlegal.com

Lucinde Rhoodie

Director:
Dispute Resolution
T +27 (0)21 405 6080
E lucinde.rhodie@cdhlegal.com

Clive Rumsey

Sector Head: Construction & Engineering
Director: Dispute Resolution
T +27 (0)11 562 1924
E clive.rumsey@cdhlegal.com

Belinda Scriba

Director:
Dispute Resolution
T +27 (0)21 405 6139
E belinda.scriba@cdhlegal.com

Tim Smit

Sector Head:
Consumer Goods, Services & Retail
Director: Dispute Resolution
T +27 (0)11 562 1085
E tim.smit@cdhlegal.com

Marelise van der Westhuizen

Director:
Dispute Resolution
T +27 (0)11 562 1208
E marelise.vanderwesthuizen@cdhlegal.com

Joe Whittle

Director:
Dispute Resolution
T +27 (0)11 562 1138
E joe.whittle@cdhlegal.com

Roy Barendse

Executive Consultant:
Dispute Resolution
T +27 (0)21 405 6177
E roy.barendse@cdhlegal.com

BBBEE STATUS: LEVEL ONE CONTRIBUTOR

Our BBBEE verification is one of several components of our transformation strategy and we continue to seek ways of improving it in a meaningful manner.

PLEASE NOTE

This information is published for general information purposes and is not intended to constitute legal advice. Specialist legal advice should always be sought in relation to any particular situation. Cliffe Dekker Hofmeyr will accept no responsibility for any actions taken or not taken on the basis of this publication.

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 cdh Stellenbosch@cdhlegal.com

©2025 14757/JUN

