

https://dspace.org/

Chatbots and negligent misstatements Commonlaw delictual liability

Emile Zitzke

https://hdl.handle.net/10539/45898
IBSS

Downloaded from a DSpace repository. https://dspace.org/

BAND 88 NOMMER 2 MEI 2025

JOURNAL OF CONTEMPORARY ROMAN-DUTCH LAW ISSN 1682-4490

THR HR

Tydskrif vir Hedendaagse Romeins-Hollandse Reg

Redakteur
Coenraad Visser B Com LLB LLM

Assistentredakteurs
Hermie Coetzee B Com LLB LLM LLD
Johann Knobel BLC LLD
Annelize Nienaber BHons LLM LLD

LexisNexis®

MEI 2025 • TYDSKRIF VIR HEDENDAAGSE ROMEINS-HOLLANDSE REG • 161–286

88(2)

BAND 88 NOMMER 2 MEI 2025

HR
HR

PRYSE viii

ARTIKELS

Separated body parts and the law of personality in South Africa – *by R Ahmed* 161

Chatbots and negligent misstatements: Common-law delictual liability – *by Emile Zitzke*..... 180

The disconnect between statutory and customary adoption in South Africa: What is the way forward? – *by Sandra Ferreira*..... 198

To be or not to be South African – income tax implications of leaving, or staying when you want to leave – *by Thabo Legwaila* 220

Recommending key principles for a reformed South African financial ombud scheme framework – *by MM Koekemoer* 234

AANTEKENING

Family dispute resolution in Japan: Lessons for South Africa? – *by Amanda Elizabeth Boniface* 251

VONNISSE

Liability for pure economic loss: The Supreme Court Of Appeal in downshifter mode – *Edward Nathan Sonnenberg Inc v Hawarden* 2024 5 SA 9 (SCA) – *by Johan Scott*..... 263

Permanence of the termination of a servitude upon merger, or its revival upon the separation of the tenements – *Turnbury House Properties (Pty) Ltd v Wallin* (18924/2020) [2022] ZAWCHC 95 (18 May 2022) – *by L-A Kiewitz* 276

Chatbots and negligent misstatements: Common-law delictual liability*

Emile Zitzke
LLB LLD
Associate professor in Law, University of the Witwatersrand, Johannesburg

OPSOMMING

Chatbots en nalatige misstellings: Gemeenregtelike deliktuele aanspreeklikheid

Klets- of geselsbots is rekenaarprogramme wat gesprekke met mense kan voer. Gewoonlik stel 'n persoon 'n vraag aan die kletsbot, waarna dié bot bepaalde inligting bewerk, en dan die persoon se vraag beantwoord. In 'n onlangse Kanadese saak, *Moffat v Air Canada* 2024 BCCRT 149, het 'n lugredery se kletsbot inkorrekte inligting aan 'n klant verskaf wat tot skade gelei het. Die Kanadese Civil Resolution Tribunal het beslis dat die lugredery aanspreeklik is vir die uitlatings van sy kletsbot, soos 'n lugredery vir misstellings wat op 'n statiese webtuiste gekommunikeer word, aanspreeklik sou wees. In hierdie artikel oorweeg ek of 'n Suid-Afrikaanse hof in 'n derglike geval 'n soortgelyke beslissing sou maak ingevolge die gemene deliktereg. Na 'n bespreking van die *Moffat* saak, word die vyf elemente van Aquiliese aanspreeklikheid op die feite toegepas: skade, die handeling, onregmatigheid, nalatigheid, en kousaliteit. Die gevolgtrekking is dat die Suid-Afrikaanse gemenereg 'n moontlike remedie vir 'n eiser verskaf in die geval van 'n nalatige misstelling deur 'n kletsbot. Litigasie oor hierdie tipe feitestel kan egter onbekostigbaar wees, veral wanneer deskundige getuienis vereis word. Daarom word 'n tentatiewe voorstel gemaak om die internasionale arena van kunsmatige intelligensie-aanspreeklikheid dop te hou vir moontlike statutêre ontwikkelings.

1 INTRODUCTION

A chatbot is a "bot which is a computer program[me] that is capable of communicating with humans, usually in a simple way".[1] More colloquially and practically, a chatbot is a computer programme that can hold a conversation with human beings (the term derives from the words *chat* and ro*bot*).[2] The conversation between the computer and the human typically unfolds through the input of human language by the user, the processing of the same by the computer programme, and outputs produced in language that the user can understand.[3] The applications of chatbots are vast.

I can flag three notable examples that already are part of many people's everyday lives to bring the discussion closer to home. In the first instance, many

* I should like to acknowledge the Oppenheimer Memorial Trust for the financial support it provided for my sabbatical research. This paper was written at the Centre for Transformative Private Law (ACT) at the University of Amsterdam, where I was a visiting fellow in 2024. I am also indebted to Leo Boonzaier who first drew my attention to the case under discussion.

1 See Ince (ed) *A dictionary of the internet* (2019).

2 See Leaua & Didu "Chatbots. Legal challenges and the EU policy approach" 2021 *Perspectives of Law and Public Administration* 210 211.

3 Ciesla *The book of chatbots: From ELIZA to ChatGPT* (2024) 11.

of us have visited websites where users find a pop-up message asking – "How can I help you today?" Users can pose a question and the computer programme then sifts through information at its disposal (largely dependent on how it is programmed by its creator) and answers the user's questions in varying degrees of complexity.[4] Secondly, on our mobile phones, we have programmes like Siri, Alexa, or Google Assistant, that are activated by voice commands (such as "Hey Alexa!"), followed by an instruction that can include asking the chatbot to provide information by conducting an internet search, or executing tasks like writing messages, operating Global Positioning Systems (GPSs), and playing music.[5] Thirdly, presently the bane of the existence of some law professors, users can ask Large Language Models (LLMs) (like ChatGPT) a question so that the computer programme sifts through huge quantities of information and summarises its findings.[6]

We might say that chatbots show the characteristics of artificial intelligence (AI) when the machine replicates human-like intelligence.[7] If we accept that human intelligence is fallible and that all people make mistakes, it is surely possible that artificial intelligence, created by humans, can also make mistakes.[8] This is not just a hypothetical musing. Recently, in Canada, a company was held liable for a negligent misstatement by a chatbot on its website.[9] To wit, no South African court has decided a case dealing with the delictual liability at common law for a negligent misstatement made by a chatbot, and few academic authors in South Africa have considered the phenomenon of chatbots.[10]

---

4 As the discussion below unfolds, it will become clear that this is the main type of chatbot under discussion here.

5 Ciesla 82–84.

6 For a specific take on teaching and learning in higher education, see Singh "Maintaining the integrity of the South African university: The impact of ChatGPT on plagiarism and scholarly writing" 2003 *SA J on Higher Education* 203 204.

7 There is no universally accepted definition of AI. In this regard, see Walters & Novak *Cyber security, artificial intelligence, data protection & law* (2021) 40 *et seq*. Some jurisdictions have started legislative processes to regulate it. In the European Union, the European Parliament and Council recently passed the Artificial Intelligence Act in April 2024. Art 3 of that Act defines an "AI system" as a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments". In the United Kingdom, the House of Lords Select Committee on Artificial Intelligence published a report *AI in the UK: Ready, willing and able* (2018) HL Paper 100. In that report, AI is defined as "technologies with the ability to perform tasks that would otherwise require human intelligence, such as visual perception, speech recognition, and language translation" (para 10) bearing in mind that modern forms of AI "have the capacity to learn or adapt to new experiences or stimuli" (para 11).

8 For an overview of recent mistakes made by AI-powered chatbots, see Brown "Bots behaving badly: A products liability approach to chatbot-generated defamation" 2023 *J of Free Speech L* 389 390–392.

9 *Moffat v Air Canada* 2024 BCCRT 149.

10 In the medico-legal context, see Donnelly "First do no harm: Legal principles regulating the future of artificial intelligence in health care in South Africa" 2022 *PER/PELJ* 1 17–21 who briefly considers civil liability implications for the situation where health care practitioners use AI, including chatbots, in their practices. For a perspective on insurance law and chatbots, see Huneberg "The future of robo-advisors in the South African insurance industry: Is the South African regulatory framework ready?" 2020 *SA Merc LJ* 175. Neither of those contributions set out to deal with the chatbot-delict interface systematically. For a perspective on American law, which is very different to ours, see Bathaee "Artificial intelligence opinion liability" 2020 *Berkeley Technology LJ* 113.

My aim in this article is to fill this gap in the literature by reflecting on how a South African court may decide a case based on a set of facts similar to that of *Moffat*, based on our common law of delict. I recognise that the Canadian law of torts and the South African common law of delict are not traditionally thought of as fertile ground of legal comparison, because the two legal systems are so different. To be clear, I will primarily use the Canadian case of *Moffat* as a springboard to illustrate a practical problem. In that process, I will necessarily provide critical reflections on some of the arguments made by the parties in the case and the legal findings made by the tribunal, but I do not aim to present a systematic and extensive legal comparison of Canadian and South African extra-contractual liability law.

This article forms part of a evolving body of scholarship around the law–AI interface in South Africa.[11] I shall limit my discussion to the common law of delict and shall not consider the possible applications of the common law of contract[12] and/or statutory consumer protection laws to a case of this nature,[13] though I imagine this could be fruitful terrain for future research. I shall start by explaining the facts and findings of the *Moffat* case as a point of departure for the discussion. This will lead me to consider the South African law of delict on negligent misstatements and its application to the issue at hand.

2 THE CANADIAN *MOFFAT* CASE

2.1 Facts

Some airlines offer discounted "bereavement fares" for passengers who are travelling on account of the imminent or recent death of a close family member. At the time of the *Moffat* decision, and at the time of writing this article, Air Canada was such an airline. The claimant in this case lost his grandmother. In relation to her death, he booked a flight with Air Canada via their website. Before making the booking, the claimant asked the airline's chatbot about bereavement fares. The chatbot provided the following response, which was captured in a screenshot:[14]

> "Air Canada offers reduced *bereavement fares* if you need to travel because of an imminent death or a death in your immediate family.... If you need to travel immediately or have already travelled and would like to submit your ticket for a reduced bereavement rate, kindly do so within 90 days of the date your ticket was issued by completing our Ticket Refund Application form."

The claimant relied on the information provided by the chatbot and booked "normal" tickets in the belief that he would be able to claim a bereavement discount retroactively. The evidence showed that bereavement fares cost around

---

11 Two recent books by South African authors deal with the AI-law interface: Marwala & Mpedi *Artificial intelligence and the law* (2024); Ncube *et al* (eds) *Artificial intelligence and the law in Africa* (2023). Neither of these works deals with the conundrum of non-contractual liability for injuries caused by AI through a systematic treatment of the South African law of delict.

12 I acknowledge that careful contracting may exclude the possibility of a delictual suit following a chatbot's negligent misstatement. However, in the absence of such intelligent use of contractual terms, a claim persists based on the common law of delict for misstatements made during the pre-contractual phase (*Bayer v Frost* 1991 4 SA 559 (A)).

13 See Consumer Protection Act 68 of 2008 (CPA). Tangentially relevant reflections on this issue are made by Eiselen "Digitisation and consumer law in South Africa" 2021 *TSAR* 436.

14 *Moffat* para 15 (original emphasis).

46 per cent of the normal ticket prices.[15] Within the 90-day period explained by the chatbot, the claimant applied for his substantial retroactive discount. Then the dispute arose: Air Canada responded to the claimant's request saying that the chatbot used "misleading words".[16] In the chatbot's response, the term "bereavement fare" was hyperlinked to the airline's bereavement travel website that contained the company's actual policy.[17] In terms of that correct policy on the website, a bereavement fare had to be applied for before the tickets were booked and could not be applied retroactively.

The claimant instituted a claim against Air Canada in the Civil Resolution Tribunal. This tribunal is mandated to resolve private-law disputes within its jurisdiction in an "accessible, speedy, economical, informal and flexible" manner.[18] According to the Tribunal's regulating statute,[19] the tribunal has jurisdiction over, among other things, small claims damages.[20] Given the informal nature of the tribunal's proceedings, the claimant here did not have to formulate his claim in the style of legal pleadings.[21] The claimant simply asked for the discount that he believed was due to him. The tribunal member who adjudicated the dispute determined that the legal issue in this case was whether the requirements for the tort of negligent misrepresentation had been met.[22]

2.2 Decision on negligent misrepresentation

The tribunal relied on the decision of the Canadian Supreme Court in *Queen v Cognos* to reiterate the requirements for the tort of negligent misrepresentation:[23] (a) a duty of care owed by the alleged wrongdoer to the claimant; (b) an untruthful, inaccurate, or otherwise misleading statement; (c) negligence; (d) reasonable reliance on the statement by the claimant; and (e) resulting damage. The tribunal's reasons for its decision, structured according to the elements from *Queen*, can hardly be compared to a court judgment. It is clearly drafted with the aim of being an accessible document that the parties can understand in lay terms, in this way "putting the public first".[24] For this reason, the application of the law to the facts in this decision is rather simplistic.

As to the first requirement, the tribunal decided that Air Canada owed a duty of care to the claimant because of the commercial relationship between them.[25] Air Canada tried to argue that the chatbot was a separate legal entity and so the airline could not be held liable for what the chatbot had said. The tribunal rejected that argument. It reasoned that the chatbot in question was simply part of the airline's website — whether the website contains static information or has an interactive component does not change the essence of a website.[26]

---

15 Paras 18–19.
16 Para 22.
17 Para 16.
18 See s 2(2)(*a*) of the Civil Resolution Tribunal Act SBC 2012 c. 25.
19 S 118(1)(*a*).
20 Small Claims Act RSBC 1996 c. 430.
21 Regarding the benefits of this user-centred dispute resolution system, see Salter & Thompson "Public-centred civil justice redesign: A case study of the British Columbia Civil Resolution Tribunal" 2016–2017 *McGill J of Dispute Resolution* 113 125–127.
22 *Moffat* para 24.
23 1993 CanLII 146 (SCC).
24 See Salter "Online dispute resolution and justice system integration: British Columbia's Civil Resolution Tribunal" 2017 *Windsor Yearbook of Access to Justice* 112 118.
25 *Moffat* para 26.
26 Para 27.

Air Canada admitted that the statement made by the chatbot was false. The second element for liability had been established by acquiescence.

The third element of a negligent breach of the duty was met, because "Air Canada did not take reasonable care to ensure its chatbot was accurate".[27] It seems that the tribunal reached this conclusion by inference.

The tribunal determined that the claimant's reliance on the statement was reasonable in the circumstances, because he obtained the information from the airline's website and could expect accuracy on any part of it, which met the fourth requirement.[28] It seems as if Air Canada argued that the claimant should have double-checked the chatbot's statement against the information about bereavement fares found on the other end of the hyperlink. The tribunal rejected that argument. It found that there was no reason for the claimant to believe that the chatbot's information required double-checking.

Finally, the claimant relied on the statement and would not have booked the normal ticket if he had known that no discount would be forthcoming after the fact.[29] Put differently, if he had received the correct information, he would have followed the correct procedure and would have obtained the discount. The tribunal awarded the claimant damages to place him in the position he would have been in, but for the negligent misrepresentation – it effectively held Air Canada retrospectively to the bereavement discount.[30]

3 REFLECTIONS FROM/FOR SOUTH AFRICA

3.1 Points of departure

In the discussion below I determine whether a South African court, judging similar facts in terms of our common law, would have come to the same conclusion. As a starting point, it should be noted that the South African common law of delict does not recognise a separate delict or tort called "negligent misrepresentation". Instead, the phenomenon of negligent misstatements is accommodated under the general principles of the common law of delict, historically rooted in the *actio legis Aquiliae*.[31] The traditionally accepted elements for Aquilian liability are conduct, wrongfulness, fault, causation, and damage. These five elements enjoy the support of the Constitutional Court[32] and many modern academic commentaries.[33]

In what follows I reflect on the elements for Aquilian liability and their implications for negligent misstatements made by a chatbot, akin to what happened in *Moffat*. I shall start with a discussion of damage, because it is the least controversial element. From there, I shall discuss conduct in some detail covering this

27 Para 28.
28 Para 29.
29 Para 30.
30 Para 33.
31 According to some scholars we could call this area of law "Aquilian liability". See, *eg,* Fagan *Aquilian liability in the South African law of delict* (2019); Boberg *Law of delict* vol 1 *Aquilian liability* (1984).
32 See *De Klerk v Minister of Police* 2021 4 SA 585 (CC) para 13.
33 See Neethling & Potgieter *Law of delict* 8 ed (2020) 4; Loubser & Midgley (eds) *The South African law of delict* 3 ed (2017) 24; Van der Walt & Midgley *Principles of delict* 4 ed (2016) para 3. Cf Fagan who contends that there are seven elements and not five, splitting damage into harm and loss, and causation into its factual and legal subcomponents. Technicalities aside, he effectively keeps the substance of the five elements in place and for the sake of tradition I will rely on the classical five-elements approach ((2019) x).

element's basic principles as well as the possible arguments that claimants and wrongdoers could raise in this regard. The element of conduct lays the foundation for the discussions of negligence, causation, and wrongfulness, which each is discussed in turn.

3.2 Damage

The element of patrimonial damage would be least controversial in this case and so I will only make brief remarks about it.[34] According to South African law, patrimonial loss is defined as the diminution of a person's estate (the totality of their rights and duties sounding in money).[35] We assess a claimant's damage by subtracting the hypothetical financial position that the claimant would have been in but for the alleged damage-causing event from the claimant's current financial position because of the alleged damage-causing event; the negative difference between the two indicates the assessed harm.[36] In a case like *Moffat*, the claimant paid $1 630.36 as the full airfare (his current position), while he would have hypothetically paid only $760 if the promise of the bereavement discount had been fulfilled (his position but for the chatbot's misstatement). Accordingly, the $870.36 difference is the patrimonial loss that the claimant had suffered.[37]

3.3 Conduct

3.3.1 *The basics*

According to the academic literature, the general element of conduct refers to the damage-causing event and requires a human to have behaved voluntarily, either through an act or an omission.[38] The term "human", in this regard, really connotes "legal subject", because all the academic commentators agree that natural and juristic persons alike are capable of conduct for purposes of the law of delict.

In South Africa, we have a strong common-law tradition of conferring legal subjectivity only on natural persons and juristic persons.[39] Natural and juristic persons are, by definition, capable of being bestowed with legal rights and capacities and bound by legal duties.[40] In the context of a delict, natural and juristic persons may have rights (the infringement of which may give rise to a delict) and duties (both to prevent delicts and to repair them when they arise). Juristic persons necessarily act through the instrumentality of natural persons, who are either its agents or employees. Practically, it makes sense to limit delictual wrongdoers to natural and juristic persons only, because only they are notionally capable of fulfilling a corrective justice obligation (for example, to pay damages) when harm has been suffered.

In terms of the law as it stands, a chatbot is not a legal subject. A chatbot has no legal rights, no legal duties, and no legal capacities. It cannot sue or be sued. Instead, natural and juristic persons may have rights and duties over a chatbot. This makes a chatbot a legal object. Specifically, because it is a computer program,

---

34 See, generally, Klopper *Damages* (2017) ch 4; Visser & Potgieter *Law of damages* 3 ed by Potgieter, Steynberg & Floyd (2017) chs 3–4.
35 See *Union Government (Minister of Railways and Harbours) v Warneke* 1911 AD 657 665.
36 See *Transnet v Sechaba Photoscan* 2005 1 SA 299 (SCA) para 15.
37 This is also how the tribunal went about calculating damages in *Moffat* para 34.
38 See Neethling & Potgieter 27; Loubser & Midgley 95; Van der Walt & Midgley para 70.
39 Boezaart *Law of persons* 7 ed (2021) 3.
40 *Ibid*.

it is intellectual property.[41] I regard this to be a clear and settled legal position at present. I would, therefore, argue that the line of argument employed by Air Canada in *Moffat* – that a chatbot is a legal entity in own right – would similarly fall on its face in South Africa.

That being said, I acknowledge that there are some international debates about whether we ought to confer legal subjectivity on AI, because some of the new technology can act autonomously to varying degrees, imitating human behaviour.[42] In *Carmichele v Minister of Safety and Security*,[43] the Constitutional Court made it clear that the judiciary's power to develop the common law applies to incremental developments while major changes would fall in the ambit of the role of the legislature. Bringing about a legal development to the effect that a chatbot is a legal subject would signify a radical alteration of the common-law tradition which would best be performed by the legislature rather than a court using its powers to develop the common law.[44]

The fact that a chatbot is not a legal subject in terms of South African law does not create an unsolvable problem for purposes of delictual liability. In the material world, we know that tangible property can be used to cause harm to others. A human actor can use a gun as an instrument to inflict harm upon another, and a victim would logically sue the human behind the gun and not the gun itself. In the cyber world, a chatbot can be deployed analogously as an instrument of harm. Without stretching the legal imagination, a victim should pursue the legal subject(s) behind the chatbot as the actor(s) for purposes of the common law of delict.

In *Moffat*, it is not entirely clear from the tribunal's reasons for its decision what conduct led to the claimant's injury. This murkiness is in part because conduct is not necessarily a constitutive element of the tort of negligence (or its sub-category of negligent misrepresentation). The tribunal in *Moffat* hints that the conduct in question is the act of making a statement to the claimant. It will be recalled that the tribunal states that when it comes to determining liability, we should see no difference between a website that contains static text and one that has a chatbot that interacts with a human. This is an interesting statement that requires further reflection. The critical question that arises for purposes of reasoning through a South African analogous case is whether Air Canada — the legal subject behind the legal object — made a statement to the claimant and, relatedly, how a claimant would go about proving that.

3.3.2 *The claimant's possible argument*

At this point it may be useful to pause and reflect on the evidentiary threshold by which the element of conduct must be proved in South African law. The claimant must establish the wrongdoer's conduct on a balance of probabilities based on available evidence, as with all civil matters.[45] Often, showing that a human actor

41 See Heaton *Law of persons* 6 ed (2021) 3.

42 The literature on this debate is vast. For a recent argument that AI may be labelled as persons, see Doomen "The artificial intelligence entity as a legal person" 2023 *Information & Communications Technology* Law 277.

43 2001 4 SA 939 (CC) para 36.

44 I also am not convinced that the legislature should do this, but that debate is outside of the scope of this article. For an European perspective, see Nevejans "European civil law rules in robotics" (2016) *Report for the European Parliament's JURI Committee* https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU (2016)571379 (accessed 17-06-2024).

45 See *Jooste NO v Minister of Police* 1975 1 SA 349 (E) 351C–E.

voluntarily did, or failed to do, something will be uncontentious and so the claimant's pleadings on conduct would be accepted as true. However, our case law has shown that there may be situations where establishing conduct may be contested from the outset. The textbook case where conduct is contested is when the defence of automatism is raised, and the alleged wrongdoer avers that voluntariness was absent.[46] However, there are other cases where conduct is challenged from the outset on account of supposedly shaky evidence linking the alleged wrongdoer to the conduct in question.

A reported case that illustrates this point played out on the internet. In *Dutch Reformed Church Vergesig v Sooknunan*,[47] the court had to determine whether the alleged wrongdoer created a website and published comments on Facebook (in other words, whether the wrongdoer acted) for purposes of an interdict to remedy a privacy violation. The alleged wrongdoer asserted a blanket denial: He neither created nor controlled the website and Facebook page, even though he was aware of it.[48] Satchwell J used inferential reasoning to find in favour of the claimant:[49] The alleged wrongdoer was a pastor of a church with the unique name of "Glory Divine World Ministries" (peculiarly abbreviated as "GDwm"), which also was the name and abbreviation used on the Facebook page; the contents on the page showed intimate personal knowledge, written in the first person, that only the pastor of GDwm would have been able to communicate; recent pictures of the pastor and his résumé were posted on the page; and the pastor's name was used and links were provided to the church's website. If the pastor knew about the page and believed that it was a rogue page impersonating him and his church, the court found it peculiar that he never tried to stop the fraudsters.[50] Overall, the probabilities weighed in favour of the claimant that the alleged wrongdoer acted for purposes of a delictual dispute.[51]

This can be applied to a *Moffat* analogue case, too. The claimant can establish conduct on a balance of probabilities by providing various facts that inferentially lead the fact finder to the alleged wrongdoer's act. Of great assistance in *Moffat* was the claimant's screenshots that were taken of the website and chatbot interaction. It would be important for a claimant to capture the website address in the browser along with the content communicated to and by the chatbot. Recent case law suggests that screenshots are, in principle, admissible in court, as long as the requirements for admissibility are met.[52] If the chatbot's communication

46 See *Molefe v Mahaeng* 1999 1 SA 562 (SCA).
47 2012 6 SA 201 (GSJ).
48 Para 33.
49 Paras 34–35.
50 Paras 38–39.
51 Paras 41.
52 Most recently, in *Sol Plaatje University v SRC of the Sol Plaatje University* (1471/2023) [2024] ZANCHC 7 (26 January 2024), the court was willing to accept, in principle, a screenshot of a WhatsApp group as evidence on condition that it met the requirements for the admissibility of hearsay evidence. In that case, the hearsay requirements were not met and so the evidence was inadmissible. In *Van Driel Boerdery Vennootskap 2004 t/a Die Groen Oase v Commissioner South African Revenue Services* (1901/2015) [2016] ZAWCHC 218 (5 January 2016), a screenshot of Google Earth was admitted as evidence because its authenticity was uncontested. A screenshot is a form of "data" ("electronic representations of information in any form") that is "electronically stored" thus comprising a "data message" for purposes of s 1 of the Electronic Communications and Transactions Act 25 of 2002. Ss 15(1)–(2) of the Act make it clear that data messages can be admissible as evidence and the various factors listed in s 15(3) would be considered in assessing the weight of the evidence.

appeared on the alleged wrongdoer's official website as captured on a screenshot, I would argue that in terms of South African law this would be sufficient to establish *prima facie* conduct similar to what was observed in *Sooknunan*: The impression created on a balance of probabilities is that Air Canada (through its agents and/or employees) made a statement to the claimant.

3.3.3 *The alleged wrongdoer's possible pleas*

After the claimant has established conduct at first glance, the alleged wrongdoing company could reply to a letter of demand or defend itself in a plea to the particulars of claim, raising various notional defences.

In *Moffat*, the wrongdoer, rather bafflingly, tried to raise a defence that we could interpret in South Africa as a defence against conduct, alleging that the chatbot was a separate legal entity (*ie* Air Canada was not the one acting), with no evidence or law supporting that claim. In South Africa, that type of defence would similarly fail because, as I have shown above, a chatbot is intellectual property and not a legal subject.

But there are three other notional defences that a company like Air Canada might be able to raise in future in South Africa.

In the first instance, if the facts are in the alleged wrongdoing company's favour, it could raise a defence similar to what was (unsuccessfully) attempted in the *Dutch Reformed Church* case, alleging that the website captured in the screenshot is not the company's website – for example, that the claimant visited a site called canadacheapflights.com instead of aircanada.com.

Secondly, the alleged wrongdoer can defend itself by saying that the chatbot's statement amounts to a type of *casus fortuitus* or *vis major*.[53] While natural disasters are unlikely to cause a chatbot to spin out of control, one could imagine a situation where a hacker plants a chatbot on a website, despite reasonable steps by the company to prevent hacking of this kind.[54] In that situation, the company might escape common-law liability if it would be too difficult to pin negligence, wrongfulness, and causation on anything that the company did or did not do. It would, of course, be an entirely different situation if the carelessness of the company made their website vulnerable to hacking, in which case their failure to secure the website may be the relevant conduct in question.

Thirdly, an alleged wrongdoer in the form of a company may raise a defence related to independent contractors being involved in the chatbot's functioning. This defence requires some careful explaining and reasoning. To develop the argument effectively, I shall start by exploring a simple material-world analogy and then add layers of complexity to it that will eventually lead to a discussion about how chatbots are made and used in the online environment.

As a start, let us consider the situation where a sole proprietor manages her own business and for some reason does so without the luxury of modern technology. If she wants to communicate information about her services or products to customers, she could advertise it by making a printed poster or print newspaper advert. In the simplest case, she would design her own marketing material and ensure the accuracy of the information provided. If the information was incorrect

---

53 I borrow this idea from s 99(2)(*a*) of the Protection of Personal Information Act 4 of 2013, being fully conscious of the fact that the classical textbooks on the common law of delict do not explore this idea in any detail.

54 On the possible malicious uses of chatbots, generally, see Ciesla ch 7.

and led others to suffer harm, her conduct of creating and publishing the material would have been the basis for the delict. The conduct would be clear: the sole proprietor communicated incorrect information to customers.

Suppose now that, in a slightly more complicated case, a company like Air Canada publishes print material (posters, pamphlets, advertisements, and/or letters) containing their bereavement policy, hoping to attract future customers. There are many ways that such marketing material could have come to the knowledge of the public. In a basic case, an authorised employee/agent of the company would create the material and distribute it as part of her core functions at the company. If she made a negligent and wrongful misstatement in the material, the company would likely be held vicariously liable for the misstatement made by her.[55] However, other actors could be involved too.

An employee/agent from Air Canada could write or type the information intended for communication and an independent contractor may have created the beautified material and published it. If the independent contractor made a misstatement contrary to what the employee/agent communicated, it cannot be said that the company made the statement, or that it should be held vicariously liable without further ado. Independent contractors are not agents or employees (read: instruments of the company) and so do not attract vicarious liability for their principals.[56] It may be a different situation if the independent contractor had a reputation for being a reckless worker, in which case the appointment of the independent contractor with the bad reputation may be a negligent act on which delictual liability might be pinned.[57] However, if a company employee or agent inspects the independent contractor's work and approves it, we are once again in vicarious liability waters, because that employee would have caused the publication of the statement that inflicted harm.

This brings us to the cyber world analogies. Once upon a time, when company websites were still solely comprised of static content, an authorised employee of Air Canada would have typed the text required on that webpage, in the course and scope of that employee's employment. If a negligent misstatement is made at this early stage of the process, vicarious liability could be imposed on the company for that employee's initial misstatement. After the initial typing of the content, a website designer could get involved. If that designer is employed in-house, then we would be dealing with potential vicarious liability once again. If the designer is an independent contractor whose work was not approved or in line with the instructions received, then the company probably did not act for purposes of the law of delict, for reasons already canvassed above.

Enter the chatbot used by Air Canada. The artificial intelligence technology behind a chatbot is complex and many people could be involved in its chain of creation and the eventual production of an erroneous statement to the public. As a starting point, it must be understood that a chatbot is programmed to operate in accordance with the instructions given to it.[58]

---

55 For a recent example of national appeal-level authority on vicarious liability in a dispute between non-state actors *inter se*, see *Stallion Security v Van Staden* 2020 1 SA 64 (SCA) para 14.

56 See *Chartaprops 16 v Silberman* 2009 1 SA 265 (SCA) para 28.

57 See *Langley Fox Building Partnership v De Valence* 1991 1 SA 1 (A) 12H–I which was applied in *Pienaar v Brown* 2010 6 SA 365 (SCA).

58 Leaua & Didu 212.

Some simple chatbots would work in an algorithmic way with very clear instructions about how to process an input into an output.[59] In the abstract, for example, if the user asks "x?" (the "input"), then the chatbot *must* source the information from a defined source (called the "domain") and *must* say "y" (the "output"). More concretely, for example, if the user asks, "can I claim bereavement retrospectively?", the chatbot might be programmed to work through Air Canada's website, and only that website, collect information on bereavement and retrospectivity, and communicate that information to the user. How the information is sourced and processed depends on the programming which depends, in part (and in theory), on what the company wants. For example, I have seen chatbots that simply reply to the question posed by providing the user with a link that leads to the appropriate page of the company's website where the user could find the information required. More elaborate programming can have the chatbot extracting a response based on defined source material and communicating that to the user, for example, replicating a relevant paragraph from a company's webpage. In an even more extravagant form, a chatbot may be programmed to consider a wider range of sources and/or to construct answers based on its own summary and/or paraphrase of a holistic reading of material available to it – such a more advanced chatbot would use patterns and probabilities observed in the sources to provide an output.[60] It seems like these more autonomous chatbots are a particular fountain from which common-law delictual liability may flow, though it must be noted that even the less advanced chatbots may end up communicating incorrect information as a result of defects in the computer programme's design, coding, and/or testing.[61]

In *Moffat*, there could have been many permutations that led to the incorrect information being received by the receiver of the misstatement.[62] The process of chatbot creation would inevitably start with an Air Canada employee-agent who would identify the source(s) of relevant information to be provided to the public. If misstated information is provided at this stage, that could be the hook for vicarious liability to be imposed on Air Canada. A designer or engineer could have conceptualised the chatbot's functioning incorrectly and contrary to Air Canada's instruction, allowing the chatbot to draw information from too many sources outside of the Air Canada website or allowing it to exercise an overly broad creative discretion. A programmer also could have acted carelessly in some way (including not testing the programme properly), or the person responsible for maintaining the website and the chatbot could have acted like an unreasonable person in the circumstances.[63] If these persons were employees or agents, it is a matter of vicarious liability. If they were independent contractors, the company might not be held responsible. However, even when dealing with independent contractors, the mere decision of a company and its functionaries to deploy an autonomous, creative-writing-inclined chatbot may be the careless conduct that we would want to label as the wrongful and negligent cause of harm. Of course, all of this must be reasoned effectively, using the extant doctrines of negligence and wrongfulness.

---

59 Ciesla 116. For an older but still relevant perspective, see Cole "Tort liability for artificial intelligence and expert systems" 1990 *Computer/LJ* 127 142–143.

60 Ciesla 12–15; Brown 392–397; Cole 149.

61 Ciesla 16; Gestner 245.

62 For an overview of parties who might be involved, see, *eg*, Gestner 243.

63 Ciesla 121–123.

Deciding who the wrongdoer is, or wrongdoers are, in the case of a chatbot making a misstatement seems to be a more intricate exercise compared to the simple publishing of a poster or a static webpage, because of the many possible actors involved. With that said, many of the underlying principles are the same, regardless of media involved. It is, therefore, difficult to make predictions in the air about the conduct on which liability will be determined. Each case would have to be decided on the facts, which would come to light only after both the claimant and alleged wrongdoer have properly stated their positions and provided relevant evidence.

It is understandable why the tribunal in *Moffat* did not embark on this elaborate exercise given the lack of convincing defences raised by the company, and the tribunal's limited jurisdiction and purpose of speedy resolution for small disputes. In South Africa, if a company deploying a chatbot fails to raise effective defences, as explained above, it is likely that the company would be held to have acted by making a statement to the claimant, based on a process of inferential reasoning.

3.4 Fault

In the preceding discussion I argued that the South African claimant in a local *Moffat* analogous case might be able to establish the element of conduct through a process of inferential reasoning. The airline's conduct in question would be "making a misstatement", assuming that no defences to the conduct element are raised effectively. The next question could be whether the alleged wrongdoer acted with fault. The relevant form of fault is probably negligence, because proving that the airline intended to cause harm to the claimant would be challenging and, realistically, very unlikely.

Negligence in the South African law of delict is classically tested by investigating the reasonable foreseeability and reasonable preventability of harm by weighing up the conduct of the alleged wrongdoer against that of the reasonable person.[64] Negligence litigation can be time-consuming and expensive, because this test is so intricate. The first question would be whether the ordinary reasonable person standard would apply, or whether the standard ought to be raised to that of a reasonable expert.[65] The second question would be whether the claimant's loss was reasonably foreseeable. If Fagan's recent and exhaustive survey of the case law is anything to go by, that is no simple task: The specific claimant must have been reasonably foreseeable;[66] some type of harm (and logically some extent of harm) must have been reasonably foreseeable;[67] and the general manner in which the harm was inflicted must have been reasonably foreseeable.[68] If the loss was reasonably foreseeable, the third question is whether the reasonable person (or expert) would have taken reasonable preventative steps considering four factors that all require the leading of further evidence: the extent of risk, gravity of consequences, utility, and cost or burden.[69] And, finally, it must be established whether the alleged wrongdoer failed to take reasonable steps.

---

64 See the classical authority of *Kruger v Coetzee* 1966 2 SA 428 (A) 430E–G.

65 For an overview of the case law, see Neethling & Potgieter 169–176; Loubser & Midgley 154; Van der Walt & Midgley paras 149–154.

66 Fagan (2019) 53–63.

67 *Idem* 66–76.

68 *Idem* 76–94.

69 These factors were famously laid down in *Ngubane v SA Transport Services* 1991 1 SA 756 (A) 776.

Most of these questions would be very difficult to answer meaningfully without knowing how the airline came to use the chatbot and what went wrong with the chatbot for it to provide incorrect information. In *Moffat*, the tribunal found that Air Canada "did not take reasonable care to ensure its chatbot was accurate".[70] This finding could have been made only through a process of inference, because no further, deeper reasoning is provided to support this finding. The question arising is whether a South African court could make a similar finding as to negligence through a process of inferential reasoning.

The most recent decision of the Supreme Court of Appeal on inferential reasoning in negligence is *Goliath v MEC for Health, Eastern Cape*.[71] It is a useful judgment that effectively summarises precedent from 1927 to 2015 – a legal position that has not materially changed since then. In *Goliath*, a woman underwent a hysterectomy, during which a swab was left behind in her abdomen, with all the ensuing discomfort associated with an infection. The claimant testified as well as the doctor who later discovered the swab. The medical staff involved in the hysterectomy decided not to testify. The claimant had no knowledge of what happened during the hysterectomy, because she was anaesthetised.[72] The doctor who removed the swab (Dr Muller) testified that if theatre procedures had properly been followed, a swab would never have been left behind in a patient's abdomen.[73]

The court in *Goliath* reiterated the first principle that the alleger is the one who must prove the element of negligence.[74] The Latin maxim *res ipsa loquitur* has often featured in the discourse on inferential findings of negligence. It means that the "facts speak for themselves". Put differently, the claimant has presented "facts which are sufficient to support and inference that a defendant was negligent and thereby to establish a prima facie case against him".[75] *Res ipsa loquitur* does not change the claimant's burden of proof, creates no presumptions, and does nothing otherwise magical or fantastical, other than affirm the basics on evidence by inference.[76] The simple question in *Goliath* was thus whether the testimonies of the claimant and the swab-removing doctor would lead the fact finder to infer negligence on the part of the medical team that performed the hysterectomy on the claimant. The following two quotes capture the court's findings:

> "[W]e simply do not know who was in attendance during the surgical procedure or whether the 'rigid regulations alluded to by Dr Muller had been followed. In particular we do not know whether there was a count of the swabs (or at the very least an attempt at one) prior to sewing up the patient. And if so, what was the level of training, and how experienced was the person to whom that task was assigned? … In this case the MEC has been sued in his capacity as the employer of all of the medical staff who at the relevant time attended to Ms Goliath during the course of the operation, at least one of whom would have had to perform the rather important task of checking and counting the swabs."[77]
>
> And,
>
> "[A]t the close of Ms Goliath's case, after both she and Dr Muller had testified, there was sufficient evidence which gave rise to an inference of negligence on the part of

70 *Moffat* para 28.
71 2015 2 SA 97 (SCA).
72 Para 13.
73 *Ibid*.
74 Para 8.
75 Para 10.
76 *Idem*.
77 Para 16.

> one or more of the medical staff in the employ of the MEC who attended to her. In that regard it is important to bear in mind that in a civil case it is not necessary for a plaintiff to prove that the inference that she asks the court to draw is the only reasonable inference; it suffices for her to convince the court that the inference that she advocates is the most readily apparent and acceptable inference from a number of possible inferences…. That being so, the MEC, in failing to adduce any evidence whatsoever, accordingly took the risk of a judgment being given against him. After all, it was open to the MEC to adduce evidence to show that while Ms Goliath was undergoing surgery, reasonable care had indeed been exercised by his employees. That he did not do. Nor, for that matter was so much as a version put during cross-examination to either Ms Goliath or Dr Muller on behalf of the MEC. Moreover, no explanation was advanced as to why the medical staff who attended to Ms Goliath were not called as witnesses. It may well be that in these circumstances an inference may be justified that the MEC feared that if one or more of them were to enter the witness box such person's evidence would expose facts unfavourable to his case."[78]

On this basis it was held that the claimant discharged the burden of proof, showing on a balance of probabilities, through inference, that the alleged wrongdoers acted negligently in causing her harm.[79] While the court did not spell it out – perhaps it should have done so – the inferences made from the claimant's evidence were that (a) the reasonable medical professionals would have reasonably foreseen that something used in the theatre could have been left behind in Ms Goliath's body; (b) the reasonable medical professionals would have taken reasonable steps, like counting swabs and instruments (which is the industry standard for mitigating the risk of grave consequences with minimal effort); (c) the medical professionals in questions failed to take those steps; and (d) thus they acted negligently.

If a case like *Moffat* occurred in South Africa, the claimant would have to present sufficient evidence that shows, at least through a process of inference, that the alleged wrongdoer acted negligently, similar to the findings made in the *Goliath* case. Expert witnesses would greatly assist a court in making factual determinations about how chatbots ought to be created and tested before they go live on websites. While I by no means profess to be an expert on the technological complexities involved in the creation of chatbots, an academic source on chatbots suggests that when the bots "hallucinate", this could be because of some failure to fine-tune the bot's operation.[80] It should be noted that there are many testing tools available to prevent this type of malfunction.[81] So it seems like proper testing of the chatbot would minimise the risk of hallucinations, although it is not entirely clear whether the risk could be eliminated completely. All that the claimant must show is that it is more likely than not that the chatbot was not carefully designed, created, and/or tested. Even though we might imagine a future where strong AI is able to act fully independently of humans, for the moment the technology is very much dependent on the human beings behind it.[82] Vladeck writes:[83]

> "Today's machines, as path-breaking as they are, all have a common feature that is critical in assessing liability. In each case, the machine functions and makes decisions in ways that can be traced directly back to the design, programming, and

78 Para 19.
79 *Ibid*.
80 Ciesla 16.
81 *Idem* 121–123.
82 *Idem* 17–18. Ciesla explains that "strong AI" that mimics the human ability to understand, show emotion, and learn by itself is not yet part of our reality.
83 Vladeck "Machines without principals: Liability rules and artificial intelligence" 2014 *Washington LR* 117 120.

> knowledge humans embedded in the machine. The human hand defines, guides, and ultimately controls the process, either directly or because of the capacity to override the machine and seize control. As sophisticated as these machines are, they are, at most, semi-autonomous. They are tools, albeit remarkably sophisticated tools, used by humans."

If expert evidence about the functioning and testing of chatbots is presented to a court, coupled with the patently wrong misrepresentation made by the chatbot in this case, on the official website of the airline, the holistic impression created could be that agents and/or employees of the airline acted negligently at some point. We must remember that the chatbot cannot wake up in a bad mood and give poor advice, become forgetful because of skipping breakfast, or lash out at a customer because it was irritable. The chatbot may have been designed carelessly to consider an overbroad domain of information in crafting its responses, possibly even at the request of the company. It may have been coded or programmed with an error. Its testing may have been done overhastily. Or it might have been a combination of all of these things, or none of them. The specific details of where the negligence could lie would necessarily depend on what evidence the claimant and defendant present to the court. In *Moffat*, the airline made no attempt at explaining why the chatbot made a misstatement. Defendants in South Africa could make a serious mistake in failing to lead their own evidence about negligence, as was observed in *Goliath*, because of the overall picture that might be painted by a claimant's evidence.

To be clear though, the mere fact that the chatbot made a mistake (in other words, "something went wrong") does not lead to an automatic conclusion of negligence in South African law.[84] The bold inference of negligence made by the tribunal in *Moffat*, reached solely on the basis of the fact that the chatbot made a misrepresentation, would probably not be made as easily by a South African court.

3.5 Causation

The element of causation contains two sub-elements that must both be established.[85] Factual causation asks whether there is any factual link at all between the claimant's harm and the alleged wrongdoer's conduct and it is traditionally tested with the *conditito sine qua non* test: If the damage would have occurred even without the alleged wrongdoer's conduct, there is no factual causation.[86] If the damage falls away when we hypothetically eliminate the alleged wrongdoer's conduct, then factual causation is established.[87] In hard cases involving concurrent or supervening causes and evidentiary gaps, resort may be had to a "increased risk test",[88] though the foundations of this test are shaky.[89] If there is such a factual link, legal causation asks whether that link is close enough such that it would be reasonable, just, and fair to hold the alleged wrongdoer responsible.[90] One or more legal causation test could give effect to justice in this regard including determining the proximate cause, the adequate cause, or the foreseeability of the chain of events leading to the harm (including the existence or not of a *novus actus interveniens*).[91]

---

84 On this principle, see *Buthelezi v Ndaba* 2013 5 SA 437 (SCA) para 15.
85 *Lee v Minister for Correctional Services* 2013 2 SA 144 (CC) para 38.
86 Para 40.
87 Para 41.
88 Para 60. On this name for the test, see Price "Factual causation after *Lee*" 2014 *SALJ* 491 492.
89 See also Fagan "Causation in the Constitutional Court: Lee v Minister of Correctional Services" 2013 *Constitutional Court R* 104.
90 *Fourway Haulage v SANRAL* 2009 2 SA 159 (SCA) para 33.
91 Para 34.

In a South African case analogous to *Moffat*, if we think away the negligent misstatement, the claimant's version is that he would have followed the correct procedure to obtain the bereavement discount from the outset. This was left uncontested by the airline and, because this seems logically likely, his version of factual causation would probably succeed. The misstatement is most closely linked to the claimant's decision to apply for the bereavement discount retroactively. There are no other events that appear to have misled the claimant. And it is objectively foreseeable that when incorrect information is communicated to prospective customers on a website, they would rely on that information. All indications weigh in favour of a finding of legal causation. The question of causation, and proximity as a consideration in legal causation in particular, may become trickier to answer when the wrongdoer points fingers towards other people in the chatbot's production chain. This did not happen in *Moffat*. In conclusion, I would argue that the airline caused the harm in this case.

3.6 Wrongfulness

The Constitutional Court has authoritative defined wrongfulness as follows:[92]

> "The wrongfulness enquiry focuses on the conduct and goes to whether the policy and legal convictions of the community, constitutionally understood, regard it as acceptable. It is based on the duty not to cause harm – indeed to respect rights – and questions the reasonableness of imposing liability."

In the preceding discussion, I argued that a South African claimant in a case like *Moffat* may be able to convince a court that the alleged wrongdoer has made a negligent misstatement. If we are indeed dealing with a negligent misstatement, South African case law provides us with some principles that give concrete expression to the vaguer notions like "policy", "acceptability", the "legal convictions of the community", and the "reasonableness of imposing liability".

The starting point is that a negligent misstatement (or any other conduct) that causes pure economic loss, is presumed to be lawful.[93] Instead, wrongfulness must be positively established by the claimant. A joint reading of appeal cases like *International Shipping v Bentley*,[94] *Mukheiber v Raath*,[95] and *Delphisure v Dippenaar*[96] reveal that relevant factors for a court to consider include (a) statutory duties imposed on the wrongdoer to provide accurate information; (b) the closeness of the relationship of trust between the wrongdoer and claimant; (c) the materiality of the misstatement; (d) whether it was obvious that the claimant would rely on the wrongdoer's statement; (e) the claimant's vulnerability to risk; (f) whether the statement was made in the course of business; (g) the wrongdoer's professional position and reputation; (h) the weight of the burden placed on the wrongdoer; and (i) the reasonableness of the claimant's reliance on the statement. Some of these factors naturally overlap, or at least strongly complement, one

92 First laid down in *Loureiro v iMvula Quality Protection* 2014 3 SA 394 (CC) para 53 and subsequently cited with approval in *Country Cloud v MEC, Department of Infrastructure Development* 2015 1 SA 1 (CC) para 21; *DE v RH* 2015 5 SA 83 (CC) para 18; *Oppelt v Department of Health, Western Cape* 2016 1 SA 325 (CC) para 51; *Mashongwa v Passenger Rail Authority of South Africa* 2016 3 SA 528 (CC) para 13.

93 *Country Cloud v MEC, Department of Infrastructure & Development* 2015 1 SA 1 (CC) para 22.

94 1990 1 SA 680 (A).

95 1999 3 SA 1065 (SCA).

96 2010 5 SA 499 (SCA).

another. In the *Moffat* analogous case, two factors weigh against wrongfulness and seven weigh in favour of it.

Outside of the CPA, airlines are not specially statutorily obliged to provide correct information on discount policies like, for example, an auditor making statements about a company's financial state. The factor labelled "vulnerability to risk" may also point in the direction away from wrongfulness: the claimant could have double-checked the correct bereavement discount policy by clicking on the relevant hyperlink in the chatbot's message and/or by calling the airline to confirm the accuracy of the chatbot's communications in this regard.

However, these two factors are not totally destructive of the claimant's case. The alleged wrongdoer and the claimant stand in a close commercial relationship to one another where the specific claimant posed a direct question to the airline's chatbot, which also indicates that the class of potential claimants is not indeterminate and there are no concerns of floodgates opening. The misstatement was material in the sense that it directly steered the claimant's behaviour towards the loss (to book the ticket immediately and wait to claim the discount later, while this would not have been done if the correct information was provided). I would argue that it is obvious that potential customers would rely on information provided on a company's website and that this reliance would be intensified when a specific question is posed and answered. It follows that the claimant's reliance on specific advice given in response to a specific question would be reasonable. If the chatbot suggested something ridiculous (such as that the claimant should take a "selfie" with the dead body before the bereavement fare will be processed), then the claimant would have a harder time showing that the reliance was reasonable. But nothing like that happened in *Moffat*. The statement was made on the airline's official website and so it was made in the course of its business, as opposed to bad advice given at a braai. The airline seems to be one of good repute and professional standing that would be in control of its customer relations. Requiring an airline to provide accurate information to its customers on a website is surely not an unduly heavy burden for a company to carry. Holistically viewed, I would argue, the case can be made that the airline's misstatement would be regarded as wrongful for purposes of the South African law of delict.

4 CONCLUDING THOUGHTS: TOWARDS A TRANSFORMATIVE APPROACH TO THE LAW-AI INTERFACE

It has been shown that chatbots can cause pure economic loss to unsuspecting claimants. This is a practical problem that might soon require the attention of a South African, similar to the ruling of the Canadian Civil Resolution Tribunal in *Moffat*. I have argued that all five elements for common-law delictual liability could be met if a chatbot makes a negligent misstatement in South Africa, on facts and pleadings analogous to *Moffat*. The dispute in *Moffat* is rather unique in that the alleged wrongdoer made no serious attempt at refuting any of the claimant's claims. This makes the claimant's case fairly easy to establish, as long as we keep our common law of delict's basic principles and the civil law of evidence in mind.

It has also been shown that matters could become murkier if the alleged wrongdoer denies responsibility and alleges that another person or persons in the chatbot's production chain ought to be held responsible instead. Matters could become even murkier if it is unclear which person ought to be held responsible for a chatbot's malfunctioning, especially as the AI technology develops towards increasing autonomy, potentially creating difficult evidentiary gaps in the causal

chain of events. Litigation delving deep into the functioning of the AI in question, possibly requiring a host of expert witnesses, may be time consuming and costly, ultimately at the risk of the claimant who already starts on the backfoot of economic injury.

In these AI-polluted waters, I would encourage us to hold onto two transformative constitutional principles, particularly (though there are surely many more). In the first instance, we must remember that the heart of the South African constitutional project has been a commitment to the flourishing of human dignity as a value and a right,[97] in terms of which humans beings are respected as beings with inherent worth and autonomy.[98] We can never allow our law to protect machines and profits derived from them above protecting vulnerable humans. Secondly, the South African constitutional experiment can work only if its provisions are translated into real-world results. In the context of injuries caused by new technology, our law ought to ensure effective and appropriate remedies for claimants, as required by section 38 of the Constitution.[99] While the common law can no doubt be a vehicle for an effective remedy when rights are violated,[100] there are legitimate concerns about the challenges that claimants may face when litigating against the deployers of AI, already alluded to above.

In light of these challenges that claimants might face, the European Union is currently engaged in a process of developing a law tentatively called the Artificial Intelligence Liability Directive.[101] The envisioned law would impose strict liability on so-called high-risk AI.[102] To protect claimants against loss caused by AI, the future law is likely to impose strict liability on those who decide to deploy AI. The law will also impose a presumption of causation, which could assist claimants who may find it challenging to prove the cause of harm in the face of evidentiary gaps.[103] Further, to assist claimants, the law may also provide specific rules on how evidence about the AI in question ought to be disclosed by alleged wrongdoers.[104] Once the European Union has passed this legislation, it may useful for South Africa to consider whether we ought to follow suit. A pre-emptive law of this kind may be an important step in ensuring that we remember what ultimately matters, striving for the continuous realisation of human dignity, and ensuring effective remedies when rights are violated. That is, in part, what a transformative constitutional approach to the future of the law-AI interface ought to involve.

97 See ss 1 and 10 of the Constitution of the Republic of South Africa, 1996.

98 To date, the most philosophically rich and legally analytical work on this right is Woolman "Dignity" in Woolman & Bishop (eds) *Constitutional Law of South Africa* (2008).

99 For a recent contribution on the importance of constitutionally-compliant legal remedies, see Zitzke "Constitutional damages in South Africa: A transformative approach" 2024 *African J of Legal Studies* 199.

100 See s 8(3) of the Constitution.

101 This legislation would dovetail the new EU Artificial Intelligence Act of 2024.

102 See Madiega "EU legislation in progress briefing: Artificial intelligence liability directive" (2023) 4 https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI-(2023)739342_EN.pdf (accessed 26-08-2024).

103 *Idem* 6.

104 *Idem* 7.

TYDSKRIF VIR HEDENDAAGSE ROMEINS-HOLLANDSE REG

Journal of Contemporary Roman-Dutch Law

Posbus 792 Durban 4001 PO Box 792 Durban 4001

17 April 2025

TO WHOM IT MAY CONCERN

I hereby confirm that the *Tydskrif vir Hedendaagse Romeins-Hollandse Reg / Journal of Contemporary Roman-Dutch Law* is a fully accredited law journal on the IBSS list.

According to the requirements of this journal, a manuscript is considered for publication only if it meets the following requirements:

- it has been recommended for publication following a double, blind peer review process;
- it is the author's own work and does not constitute plagiarism; and
- it has not been submitted for publication elsewhere.

The following contribution met these requirements and was published in 2025:

- E Zitzke "Chatbots and negligent misstatements: Common-law delictual liability" (2025) 88 *THRHR* 180–197.

No page fees were charged for the publication of this article.



Professor Coenraad Visser
EDITOR