

When Algorithms Mislead: Refining Situational Crisis Communication Theory for AI-Mediated Accountability

Daniel I. Nwogwugwu

Department of Communication Studies,
Northeastern University, Boston, USA

Abstract: The integration of Artificial Intelligence (AI) into stakeholder-facing organizational and crisis communication has created a crisis cluster where the immediate crisis cause is a non-human agent. This has created accountability ambiguities which current crisis communication frameworks do not fully address. This study examines *Moffatt v. Air Canada* 2024 BCCRT 149, the first publicly reported Canadian tribunal ruling to assign organizational liability for misleading information by a customer-facing AI chatbot. Drawing on data from the full Tribunal record, Air Canada's responses and a content analysis of 47 news articles, the study traces a three-phase crisis lifecycle and identifies a mismatch between Air Canada's deflection-based response and the rebuild suggestion of the Situational Crisis Communication Theory (SCCT) for high attribution crises. Building on Kelly's (2001) stewardship tradition in public relations and Prah and Goh's (2021) research on organizational responses to AI failures, this study introduces the Algorithmic Stewardship Accountability Framework (ASAF) as a refinement of SCCT's build strategy. ASAF retains SCCT's attribution mechanism, traces algorithmic output failures within the preventable cluster as a human-error product harm, and strengthens the rebuilding strategy by including a stewardship declaration comprising categorical ownership, governance disclosures, and ongoing reporting.

Keywords — Artificial Intelligence (AI), algorithmic accountability, Air Canada, crisis communication, SCCT, stewardship

SUGGESTED CITATION: Nwogwugwu, D. I. (2026). When algorithms mislead: Refining situational crisis communication theory for AI-mediated accountability. *Proceedings of the International Crisis and Risk Communication Conference, USA, 14(1)*. <https://www.doi.org/10.69931/001c.164607>

INTRODUCTION

The adoption of Artificial Intelligence (AI) has dominated several scholarly discourses since they became very popular within the last decade. In organizational communication, industry statistics reveal that “79% of organizations have embedded them into their customer experience” [1]. These AI tools are used to strengthen operational efficiency, scalability, prompt responses and promising rapid information delivery [2]. However, this adoption has outpaced the capacity of existing crisis communication theories. Where previous scholarship explored AI primarily as tools for crisis management, for example, in media sentiment or scaling volume responses [2,3], a new dynamic has emerged with implications for crisis communication and management. Recent incidents suggest that delegating high-stakes communication to AI can introduce new risks to stakeholders. When a customer-facing chatbot disseminates misinformation that leads to legal disputes and reputational harm, the question of accountability becomes structurally ambiguous. Responsibility may plausibly diffuse to the developers of the system, deploying organization, third-party vendor, training team, or the algorithm itself. This is the case with Air Canada's AI-chatbot incident [4], which offers a well-documented entry into understanding and seeking solutions to this problem.

In November 2022, an Air Canada customer, Jake Moffatt, contacted Air Canada's customer-facing chatbot regarding bereavement fares following the death of his grandmother. The chatbot responded that bereavement-rate refunds could be claimed retroactively, which, unknown to Moffatt, was misleading and directly contrasted with Air Canada's published policy. After travelling and submitting a refund claim, Moffatt was denied reimbursement, leading to legal issues. He filed a

ISSN: 3068-6539

© 2026 Copyright is held by the owner/author(s).

Publication rights are licensed to ICRCA.

<https://doi.org/10.69931/001c.164607>

claim with the British Columbia Civil Resolution Tribunal (CRT) and in their defense, Air Canada argued that the chatbot “was a separate legal entity responsible for its own actions” and Moffatt “should have verified the information against their official policy” [4]. The Tribunal rejected these arguments, ruling that the airline was responsible for its chatbot's negligence and ordered Air Canada to pay C\$812.02, comprising damages, interest, and tribunal fees [4]. The Tribunal further emphasized that customers had no reason to assume that one part of an Airline's website was more authoritative than another, as both were official channels.

This case is significant because it provides the first legally adjudicated test of an organizational deflection strategy that [2] documented but had never been subjected to tribunal scrutiny. Hence, this study traces the crisis lifecycle, examines Air Canada's response against SCCT's recommendations and proposes a refinement of the SCCT, the Algorithmic Stewardship Accountability Framework (ASAF), which expands on the rebuild posture with a stewardship declaration. The framework is offered for further empirical testing rather than as an established extension. The next section provides a brief review of the literature, methodology, and empirical findings, and extends the theoretical framework from the crisis.

LITERATURE REVIEW

AI IN CRISIS COMMUNICATION

Scholarship on AI in organizational and crisis communication has generated two opposite dimensions. The dominant dimension treats AI as a crisis management tool. For example, [5] experimental research found that chatbots can outperform human spokespersons in delivering timely (instructive) information when perceived competence was high during unresolved crises. However, trust in AI did not differ significantly from trust in human representatives. These findings show that AI was a visible crisis communicator under bounded conditions but say little about what happens when the tool becomes the source of harm. The second dimension examines AI as a crisis trigger rather than a tool. The most directly relevant precedent is [2], which analyzed 23 publicly documented AI failures, the organizational responses and the resulting media coverage. Their findings show three things. 1. They identified organization-AI-user responsibility triangle, where corporate accountability is asymmetrically maximized. Here, stakeholders consistently locate responsibility with the deploying organization rather than with the algorithm or user. 2. They explained that several conventional SCCT strategies remained effective for AI-caused crises, thereby suggesting that the framework is partly manageable rather than totally inadequate. 3. They identified the emergence of what they termed a “mirror strategy,” where organizations exploit AI accountability to deflect blame. This present study extends [2]'s work by analyzing the first legally adjudicated case in which such deflection was tested, and by specifying the response architecture that organizations should adopt when the deflection strategy is known.

Also, [3] systematic review of 177 SSCI-indexed studies on crisis communication and AI confirms the structural imbalance identified by [2]. A further consideration is that while hallucination has become common for AI-generated misinformation [6], much corporate chatbot deployment from 2020 to 2023 relied on retrieval-based systems whose failures stemmed from outdated knowledge bases and inadequate oversight rather than probabilistic generation. Thus, the Air Canada chatbot's architecture was not established at trial, and this study treats the failure as an algorithmic output failure, which is uncertain to the current underlying frameworks.

SCCT AND ITS LIMITS IN AI-CRISIS COMMUNICATION: THE STEWARDSHIP TRADITION

The Situational Crisis Communication Theory (SCCT) developed by [7] and refined across subsequent studies, remains the most dominant framework for understanding crisis responsibility and provides evidence-based suggestions on matching organizational responses to levels of attributions of stakeholder responsibility. SCCT categorizes crises into victim, accidental and preventable, based on stakeholders' attributions of organizational responsibility and suggests corresponding response strategies to each crisis cluster, ranging from denial to diminish and rebuild strategies.

Although several studies have explored the limitations of the SCCT [8, 9], a new dimension of these limitations has emerged. Although the framework has proven useful, its attribution mechanism points towards human communicative agency. However, AI-mediated communication strains these assumptions in two ways. First, the communicator is non-human, thus creating what [2] call accountability uncertainty. Second, the deployment architecture is often multi-party, as most organizations procure chatbots from third-party vendors, introducing intermediaries between the stakeholder-facing output and the system's training configurations.

Another tradition is relevant here and less used in AI-crisis scholarship. This involves organizational stewardship, introduced by [10] as the fourth stage of the public relations process, comprising reciprocity, responsibility, reporting and relationship nurturing. Stewardship offers a near-perfect framework for AI-crisis communication. This involves the recognition that organizational responsibility for stakeholder-facing systems endures beyond the immediate incident, and that maintaining trust after an AI failure demands disclosures of the exact errors and governance architecture.

METHODOLOGY

This revelatory single case study employed a content analysis of legal and media documents. [11] explains a revelatory case as one that allows researchers to observe and analyze a phenomenon previously inaccessible to scientific investigation.

Thus, this case qualifies as such because the Tribunal ruling exposed the legal, communicative and reputational dimensions of AI accountability absent in previous instances. The crisis agent is non-human; the organization attempted to shift the blame to an algorithm, and stakeholder reactions are documented in legal and media records.

Data comprised the full Tribunal decision, Air Canada's defense statements, Moffatt's evidence, the chatbot transcript reproduced in the ruling, and a corpus of 47 English-language news articles. Articles were retrieved through Factiva and Nexis Uni using the search terms "Air Canada chatbot," "Moffatt Air Canada," "Air Canada and Jake Moffatt," and "Air Canada AI chatbots," from November 2022 to June 2024. Inclusion required substantive coverage of the incident beyond ordinary mention. Outlets included *The Guardian*, *The Washington Post*, *Reuters*, BBC News, CBC News, and trade publications. Articles were coded for their tone (positive, negative and neutral), thematic framing (inductively derived) and crisis responses characterization (aligning with SCCT's deny, diminish and rebuild categories). Two coders trained on a shared codebook independently coded a randomly selected 25% reliability subsample (n=12 articles). Cohen's kappa (κ) was 0.81 for tone and 0.77 for thematic framing. Disagreements were resolved through discussions aligning with the codebook.

RESULTS

Results revealed that the crisis unfolded over two years, over three distinct life-cycle phases, each of which raised questions about AI-mediated crisis communication. The first phase was the algorithmic failure, from November 2022, when Jake Moffatt contacted Air Canada's website following his grandmother's death to request information on bereavement fare discounts and received responses from Air Canada's AI chatbot. The chatbot provided information that a retroactive refund application could be submitted within 90 days of the ticket purchase, directly contradicting the airline's published bereavement policy. Moffatt purchased a round-trip ticket for approximately C\$1,630 and submitted a refund application with the required documentation, in line with the information provided by the chatbot. The architecture of the chatbot was not established in the Tribunal proceedings, as there was no evidence on its design, training, or maintenance. Hence, this study treats this incident as an algorithmic failure, a category in which we are unsure whether the system was generative, retrieval-based, or rule-based. What is established is that the organization lacked sufficient monitoring to correct the anomaly.

The second phase involved the organization's early response in 2023, where they rejected Moffatt's refund claims and citing the actual refund policy. However, when Moffatt presented evidence from the chatbot transcript, an Air Canada representative acknowledged the misinformation and indicated that the system would be updated. The airline offered a C\$200 goodwill coupon, which Moffatt declined [4, parag 27]. This initial response is significant as it demonstrates the organization's internal awareness of the failure, but chose not to remediate the specific stakeholder harm. Additionally, the goodwill coupon served as partial bolstering, signalling an intent to maintain the relationship rather than a full rebuild without accepting responsibility. The absence of a stewardship declaration here led to the matter being taken to legal adjudication.

The third phase involved the legal battles and public crisis where Moffatt filed a lawsuit with the British Columbia Civil Resolution Tribunal against Air Canada in late 2023 up until February 2024 when judgement was passed. Air Canada's submission that the chatbot was a "separate legal entity responsible for its own actions, independent of the airline" was remarkable. The Tribunal rejected this argument, stating that Air Canada owed a duty of care to customers using its website for information and that Moffatt's reliance on the chatbot on its website was reasonable. The company was ordered to pay C\$812.02 in damages and interest to Moffatt. This ruling received wide international coverage within days, transforming a supposed minor consumer dispute into a major precedent on AI accountability.

MEDIA FRAMING AND SCCT RESPONSE MISMATCH TAKEHOLDER ATTRIBUTION

The results revealed a predominant negative tone as 34 out of the 47 articles (72.3%) were negative, 11 (23.4%) were neutral and 2 (4.3%) were positive. The negative framing focused less on the algorithmic failure than on Air Canada's response, hence, suggesting stakeholder attribution intensified after the organizational response became public. This pattern supports [2] finding that AI accountability ambiguity invites deflection that amplifies harm when challenged. Furthermore, three thematic frames emerged. First, AI governance deficit, with 35 (74.5%) articles portraying the incident as evidence of inadequate corporate AI oversight. The second theme was a response to illegitimacy, with 32 (68.1%) articles criticizing Air Canada's deflection, with coverage language such as "absurd" and "evasive" alongside the Tribunal's "remarkable" description. The final theme was precedent significance, evident in 29 (61.7%) articles, which framed the Tribunal's ruling as a turning point for AI accountability.

Applying SCCT to this crisis case reveals a distinct strategy mismatch. The algorithmic failure constitutes an internal operational failure within the organization's control, thereby placing the crisis within SCCT's preventable cluster as a human-error product harm. Thus, [7] recommends a rebuild strategy that includes apology, compensation and corrective actions. However, Air Canada's response combined bolstering (the goodwill coupon) with scapegoating (blaming a third party). Scapegoating is recommended for victim-cluster crises and not preventable crises. SCCT's guidelines predict that, when applied to high-responsibility crises, they would only amplify reputational harm because the denial is perceived as insincere, generating more outrage beyond the original issue [7]. Hence, the response illegitimacy frame corroborates this prediction.

THEORETICAL REFINEMENT: THE ALGORITHMIC STEWARDSHIP ACCOUNTABILITY FRAMEWORK

Based on empirical findings and theoretical analysis, this study proposes the Algorithmic Stewardship Accountability Framework (ASAF) as a refinement of SCCT's rebuild posture for cases where the communicative agent is an

organizationally-deployed AI system. There are three reasons for this refinement argument.

First, SCCT's cluster taxonomy is built on stakeholder attribution rather than the status of the causal agent. The Air Canada case demonstrates this, as the media and Tribunal attribute responsibility solely to the organization, thereby making it a preventable case. Secondly, the case fits [7] existing sub-classification of human-error product harm within the preventable cluster. The chatbot is a product of the organization, either developed or procured. Its failure traces human decisions about training data, content management, vendor selection and monitoring. Hence, treating algorithmic output failures as a species of human-error product harm preserves SCCT's attribution logic while also accommodating the operational features of AI-mediated communication. Lastly, the primary problem is not where AI failures fit in the cluster taxonomy. Rather, the argument is about how the rebuild strategy must be detailed when accountability ambiguity gives organizations a deflection route that human communicator crises do not afford.

Thus, ASAF proposes the stewardship declaration as the augmenting component. Drawing on [10] study, there are 3 commitments layered into rebuild. Categorical ownership is the explicit acceptance of responsibility for stakeholder-facing AI output, with explicit rejection of AI as independent agent and vendor as responsible party deflections, aligning with [2] identifications. Secondly, governance disclosure is the transparent reporting of what failed, why it failed and corrections undertaken, which extends SCCT's corrective-action component. Lastly, ongoing reporting is a commitment to continuous disclosure of AI governance practices as part of stakeholder relationships.

A procedural distinction operates within the declaration depending on whether the AI system was developed internally or externally. This distinction does not modify the level of accountability but shapes the content of governance disclosures. Table 1 summarizes the ASAF.

Table 1: Stewardship Declaration Components by Deployment Scenario

Declaration Component	Direct Deployment (self-developed AI)	Delegated Deployment (third-party AI)
Categorical Ownership	Explicit organizational acceptance of responsibility for system output; rejection of AI-as-independent-agent framing	Explicit organizational acceptance of responsibility for system output; rejection of AI-as-independent-agent and vendor-as-responsible-party framings
Governance Disclosure	Transparent account of failures in internal design, training data, testing protocols, and monitoring.	Transparent account of failures in vendor system performance and organization's procurement oversight, quality assurance, and monitoring of that system
Ongoing Reporting	Continued disclosure of changes to internal AI development, testing, and monitoring architecture	Continued disclosure of changes to vendor selection criteria, contractual oversight, and quality assurance auditing

CONCLUSION

The Air Canada case reveals an operational vulnerability in SCCT when applied to AI-mediated crises. However, ASAF tackles this by refining the rebuild strategy with a stewardship declaration designed to foreclose deflection at the level of organizational communication rather than at the level of legal liability. The framework is presented for testing as multi-cases and experiments, for example, are required to establish its predictive validity across AI-mediated crisis types, deployment architectures, cultural contexts and organizational structures. As AI systems become embedded in stakeholder-facing communication, so must crisis theories accommodate non-human communicative agents without abandoning the attribution mechanism of the SCCT.

Author Biography

Daniel Nwogwugwu holds a Postdoctoral Teaching position in the Department of Communication Studies at Northeastern University, Boston.

REFERENCES

- [1] CMSWire. (2025). State of Digital Customer Experience Report. <https://www2.cmswire.com/cp-smg-dcx-report-cx.html>
- [2] Prah, A., & Goh, W. W. P. (2021). "Rogue machines" and crisis communication: When AI fails, how do companies publicly respond? *Public Relations Review*, 47(4), 102077. <https://doi.org/10.1016/j.pubrev.2021.102077>
- [3] Cheng, Y., Li, J., Lee, J., & Liu, Y. (2025). How is artificial intelligence shaping crisis communication? A

systematic review and future research agenda. *Journal of Risk Research*, 1–21.

<https://doi.org/10.1080/13669877.2025.2579317>

- [4] *Moffatt v. Air Canada*, 2024 BCCRT 149. British Columbia Civil Resolution Tribunal. <https://canlii.ca/t/k2spq>
- [5] Xiao, X., & Yu, S. (2025). Can ChatGPT replace humans in crisis communication? The effects of AI-mediated crisis communication on stakeholder satisfaction and responsibility attribution. *International Journal of Information Management*, 80, 102835. <https://doi.org/10.1016/j.ijinfomgt.2024.102835>
- [6] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- [7] Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of Situational Crisis Communication Theory. *Corporate Reputation Review*, 10(3), 163–176. <https://doi.org/10.1057/palgrave.crr.1550049>
- [8] Barkley, K. (2020). Does one size fit all? The applicability of situational crisis communication theory in the Japanese context. *Public Relations Review*, 46(3), 101911. <https://doi.org/10.1016/j.pubrev.2020.101911>
- [9] Liu, Y., Gao, D., & Duan, H. (2025). Choosing the right strategies: An analysis of crisis response strategies in Chinese universities. *PLoS One*, 8;20(4), 1-20. <https://doi.org/10.1371/journal.pone.0321777>
- [10] Kelly, K. S. (2001). Stewardship: The fifth step in the public relations process. In R. L. Heath (Ed.), *Handbook of public relations* (pp. 279–289). Sage.
- [11] Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.