### Beyond Just Apologies: The Role of Ethic of Care Messaging in AI Crisis Communication Sera Choi

#### Introduction

Artificial intelligence (AI) is transforming industries and society, bringing both opportunities and risks. As organizations increasingly adopt AI systems, ethical considerations around transparency, bias, and accountability have moved to the forefront (Floridi et al., 2018). Failures of AI systems can quickly escalate into organizational crises, necessitating careful crisis communication to maintain relationship with publics. However, research on crisis communication strategies for AI failure crisis is sparse. Therefore, this study examines the intersection of crisis response strategies and an ethic of care message in the context of AI failure crisis.

Simply put, AI refers to software algorithms that mimic human cognitive functions such as learning and problem-solving (Russell & Norvig, 2016). With recent advances in deep learning and neural networks, AI can now perform complex tasks like facial recognition (Kusumah et al., 2022) and language translation (Cao, 2022). However, AI systems aren't perfect. For example, biased or flawed training data, algorithmic errors, and misaligned incentives can cause AI failures with harmful consequences (Amodei et al., 2016).

Crisis communication focuses on communicating information flows to mitigate reputation damage from unexpected events (Coombs, 2019). The situational crisis communication theory (SCCT) matches crisis response strategies to crisis types based on attribution of responsibility (Coombs, 2007). While SCCT (Coombs, 2007) addresses technical errors as one category of crisis (i.e., accidental crisis cluster), AI failures present distinct challenges warranting specialized analysis. From an SCCT perspective, technical malfunctions imply mechanical or software flaws disrupting business operations (Coombs & Holladay, 2002). However, undesired AI behaviors often stem not from simple technical defects, but rather complex socio-technical risks around ethics, accountability, and transparency (Neff & Nagy, 2016). For instance, biased algorithms or flawed training data can produce discriminatory AI decisions that appear technically sound but have harmful social impacts (Amodei et al., 2016).

A number of well-known incidents have shown that crisis that can unfold when companies face AI failure. In 2016, Microsoft launched Tay, an AI chatbot designed to converse on Twitter. However, within 24 hours, Tay began tweeting offensive content it learned from users. Microsoft quickly took Tay offline, however, the incident damaged its reputation (Neff & Nagy, 2016). Drawing on a different case, in 2018, Amazon employed an AI recruiting tool which, troublingly, was found to downgrade women's resumes, despite its promise to bolster diversity (Dastin, 2018). These cases show the pressing need to understand how organizations can effectively manage AI failure crises. Further research should empirically examine how strategies like denial (e.g., deny, scapegoat), diminishment (e.g., excuse, justification), and rebuilding (e.g., apology, compensation) (Coombs, 2007) differentially impact public perceptions on AI crises. Thus, the current study investigates the following research question:

RQ: In the context of AI failure crises, do response strategies (denial vs. excuse vs. apology) differ in their effectiveness on (a) organizational reputation and (b) supportive behavioral intention?

In addition to crisis response strategies, the concept of the ethic of care offers a unique lens to examine AI crisis communication. The ethic of care emphasizes moral responsibility above rigid rules, acknowledging the interdependence between individuals and their context-specific needs (Gilligan, 1993). Messages imbued with empathy, concern, and compassion boost both the credibility of the communication and its source (Seeger, 2006). In fact, rather than solely focusing on strict legalities, communication rooted in an ethic of care gives priority to moral accountability (Simola, 2003; Tao & Kim, 2017). This approach not only nurtures trust but also fosters meaningful relationships between publics and organizations.

Past studies provide a basis to hypothesize that integrating high ethic of care messages into crisis response strategy will improve organizational reputation and relationship outcomes compared to low ethic of care messages. Research on effective crisis management emphasizes the importance of expressing empathy, compassion and concern, which are the core values of ethic of care, for those affected (Seeger, 2006; Ulmer et al., 2022). Experiments found that crisis responses arising empathy increased reputation repair from crisis (Schoofs et al., 2019). Therefore, infusing high ethic of care message into crisis responses should strengthen perceived relationship quality such as organizational reputation and supportive behavioral intentions compared to low ethic of care messages, by signaling an organization's benevolent motivations and prioritization of public needs over self-interest. This aligns with Simola's (2003) model of a caring corporation managing stakeholder connections through ethic (ethic of justice and ethic of care) centered communication. Thus, the following hypotheses are suggested:

H1: High level of ethic of care message will be associated with greater levels of (a) organizational reputation, and (b) supportive behavioral intention than low level of ethic of care message.

H2: Response strategies combining with high level of ethic of care message will be associated with greater levels of (a) organizational reputation, and (b) supportive behavioral intention than response strategies with low level of ethic of care message.

#### Method

To address the RQ and hypotheses, this study conducted two experiments manipulating crisis response strategies and ethic of care message under AI failure crisis scenarios. The studies measured impacts on post crisis organizational reputation, and intention to support the organization. Study 1 investigates the effects of three strategies—denial, excuse, and apology— to an AI gender bias crisis on organizational reputation and supportive behavioral intention. Study 2 delves into the persuasiveness of these strategies when paired with varying levels of ethic of care messages.

#### **Procedures and participants**

Study 1 utilized an online survey experiment administered through Qualtrics survey software. Participants were recruited from Amazon Mechanical Turk (MTurk) and a total of 36 participants located in the United States successfully completed the study procedures. Each participant was randomly assigned to view one of three AI crisis response scenarios. After viewing the scenario, participants answered questions regarding their reactions to the organization's response.

Study 2 employed a 3 (crisis response strategy: denial, excuse, apology) X 2 (care ethics: high, low) between-subjects experimental design. Participants were again recruited through Amazon MTurk. Out of the 144 participants initially recruited for the study, seven were excluded from subsequent analyses. These exclusions were based on their failure to pass an attention check embedded within the survey. Consequently, the final sample used for data analysis consisted of 137 participants. After participants completed the survey, the debrief message was shown to the screen.

#### Stimuli

In Study 1, we created fictional AI-driven crisis scenarios based on real incidents, using a fictitious company, "Hexxa." This controlled approach ensured consistent exposure to the same scenario; avoiding biases linked to known brands. In this case, Hexxa's AI recruiting tool displayed male candidate bias. Participants reviewed three response strategies from Hexxa: a denial where they claimed non-use of the AI for recruiting, an excuse asserting the results didn't echo their views, and an apology where they acknowledged the problem and promised improvement (see Appendix A).

In study 2, the crisis scenarios and response strategies were adapted from Study 1. The three crisis responses (denial, excuse, apology) were crossed with two levels of ethic of care message (high vs. low). The high ethic of care responses added language emphasizing concern for stakeholders, relationship nurturing, use of action plan to prevent conflicts, what is right for victims rather than focusing on legalities, commitment to fostering a divers and inclusive work environment, and avoiding harm (Tao & Kim, 2017). The high ethic of care responses used more generic phrasing about understanding feelings and importance of nurturing relationships among publics (see Appendix B & C). After viewing one randomly assigned scenario, participants completed the same outcome measures as Study 1.

#### Measures

**Organizational Reputation.** A 5-item organizational reputation scale based on the measurement developed by Walsh and Beatty (2007) was used.

*Supportive Behavioral Intention (SBI).* From the previous research (Overton et al., 2021; Zeithaml et al., 1996), SBI was measured using a 3-item.

#### Results

#### **Manipulation Checks**

To assess whether the crisis response strategy manipulations were successful, participants responded to an item asking their level of agreement that the organization took responsibility for the crisis for both study 1 and study 2. A one-way ANOVA determined there were significant differences in perceived responsibility based on condition, F(2, 33) = 31.62, p < .001. Post-hoc tests revealed the deny condition (M = 2.08, SD = .79) had significantly lower responsibility perceptions than the excuse (M = 3.92, SD = 1.16) and apology conditions (M = 6.25, SD = .87) at p < .05. The excuse and apology conditions also significantly differed from each other. These results confirm the crisis response manipulations effectively portrayed varying levels of accepting responsibility through the deny, excuse, and apology strategies.

To ensure that the ethic of care manipulation was perceived as intended, a 6-items was asked (see Appendix D) expanded upon Tao and Kim (2017). A *t*-test was conducted comparing the high ethic of care condition (M = 5.62, SD = .96) to the low ethic of care condition (M = 3.21, SD = .83). The results confirmed that the high ethic of care messages was perceived as higher than the low ethic of care messages, t(137) = 14.92, p < .001.

#### **RQ: Effectiveness of Response Strategies on Organizational Reputation and Supportive BI**

A Multivariate Analysis of Variance (MANOVA) was conducted to assess the multivariate effects of the different response strategies (deny, excuse, apology) on (a) organizational reputation and (b) supportive behavioral intention.

A significant effect was observed for the response strategies on organizational reputation,  $F(2, 33) = 8.42, p < .001, \eta^2 = .338$ . Post-hoc comparisons indicated that the apology strategy was significantly more effective than the deny strategy in enhancing organizational reputation (p < .05). The excuse strategy was also found to be more effective than the deny strategy (p < .05). There was no significant difference between the apology and excuse strategies (p = .784).

There was a significant effect of the response strategies on supportive behavioral intention, F(2, 33) = 7.89, p = .001,  $\eta^2 = .324$ . Post-hoc tests revealed that the apology strategy led to higher levels of supportive behavioral intention compared to the deny strategy (p < 0.05). Similarly, the excuse strategy also resulted in more favorable behavioral intentions than the deny strategy (p < .05). As with organizational reputation, there was no significant difference between the effects of the apology and excuse strategies on supportive behavioral intention (p = .812).

# H1: Effectiveness of Ethic of Care Message on Organizational Reputation and Supportive BI

To assess the impact of the different level of ethic of care message on organizational reputation and supportive behavioral intention, independent t-tests were performed.

For organizational reputation, results indicated that participants exposed to the high level of ethic of care message (M = 5.31, SD = 1.08) rated the organization's reputation significantly higher than those exposed to the low level of ethic of care message (M = 4.75, SD = 1.12), t(137) = 3.72, p < .001.

Regarding supportive behavioral intention, participants who viewed the high ethic of care message (M = 5.48, SD = 1.06) expressed significantly higher intentions to support the organization compared to those who saw the low ethic of care message (M = 4.89, SD = 1.11), t(137) = 4.01, p < .001. Thus, H1 was supported.

# H2: Effectiveness of Response Strategies and Ethic of Care Message on Organizational Reputation and Supportive BI

To test Hypothesis 2, a 3 (response strategy: deny, excuse, apology) x 2 (ethic of care message: high, low) ANOVA was conducted for both organizational reputation and supportive behavioral intention.

For organizational reputation, there was a main effect for ethic of care message, with high ethic of care consistently outperforming low ethic of care across all response strategies: apology  $(M_{\rm high}=5.72, SD_{\rm high}=0.83 \text{ vs. } M_{\rm low}=4.88, SD_{\rm low}=0.91, p < .001)$ , excuse  $(M_{\rm high}=5.34, SD_{\rm high}=0.79 \text{ vs. } M_{\rm low}=4.42, SD_{\rm low}=0.86, p < .01)$ , and deny  $(M_{\rm high}=4.95, SD_{\rm high}=0.85 \text{ vs. } M_{\rm low}=4.02, SD_{\rm low}=0.80, p < .05)$  (see Appendix E).

A similar pattern emerged for supportive behavioral intention. High ethic of care was superior to low ethic of care across the board: apology (M\_high= 5.58, SD\_high= 0.78 vs. M\_low= 4.70, SD\_low= 0.82, p < .001), excuse (M\_high= 5.20, SD\_high= 0.75 vs. M\_low= 4.28, SD\_low= 0.81, p < .01), and deny (M\_high= 4.82, SD\_high= 0.79 vs. M\_low= 3.90, SD\_low= 0.77, p < .05) (see Appendix F). Thus, Hypothesis 2 was supported across all strategies.

#### Discussion

The primary objective of Study 1 was to explore the differences in the effectiveness of response strategies (denial, excuse, apology) in the context of AI failure crises, particularly in terms of organizational reputation and supportive behavioral intention.

The findings suggest that, in the aftermath of an AI failure crisis, response strategies play a crucial role in influencing public perceptions and intentions towards the organization. Notably, both the apology and excuse strategies were more effective than the deny strategy in enhancing organizational reputation and fostering supportive behavioral intentions. These results align with the crisis communication literature that emphasizes the value of accountability and responsibility during crisis management (Coombs, 2007). This suggests a need to refine or expand models like the SCCT to account for technology-specific crises.

The non-significant difference between the apology and excuse strategies, in terms of their impact on organizational reputation and behavioral intentions, is particularly intriguing. It suggests that, following AI failure, acknowledging the issue—whether by offering an apology or providing

an excuse—can be beneficial for organizations. This might be indicative of the public's desire for transparency and acknowledgment over outright denial in the increasingly AI-integrated landscape.

While the deny strategy was the least effective among the three, it's worth noting that there might be specific contexts or situations where denial could be appropriate, especially if the organization is falsely accused. However, in the wake of genuine AI-driven errors, our results underscore the drawbacks of using denial as the primary response strategy.

Turning to Hypothesis 1, the findings of the study supported that a high level of ethic of care message supportively influences both organizational reputation and supportive behavioral intention. This emphasizes the importance of ethical responsibility in crisis communication, especially in contexts such as AI failures, regardless of the specific response strategy employed. This aligns with the growing emphasis on ethical considerations in crisis communication research (Bowen & Coombs, 2020).

Hypothesis 2 extended the inquiry by examining how different response strategies interact with varying levels of ethic of care. The findings revealed that high ethic of care messages consistently led to more favorable outcomes across all response strategies—deny, excuse, and apology—in terms of both organizational reputation and supportive behavioral intention. This suggests that simply acknowledging isn't sufficient. Reinforcing ethical responsibility and offering clear action plans to address and investigate issues adds another layer of trust and assurance for the public. Importantly, this also implies that integrating an ethic of care element can even elevate the outcomes of the typically less effective deny strategy.

#### **Overall Implications and Conclusion**

The study provides valuable insights into the effectiveness of different crisis response strategies in the context of AI failure crises. The clear advantage of apology and excuse over denial offers practical guidelines for crisis management. Ethical considerations, specifically a high ethic of care, universally improve outcomes, irrespective of the response strategy used. Future research could explore more nuanced interactions between these variables or extend this study to other types of crises. Given the significant impact of AI on various sectors, understanding how to best manage crises in this new landscape becomes increasingly critical.

*Limitations.* This study, while shedding light on critical aspects of crisis communication in AI failures, possesses several limitations. One key limitation is the generalizability of the findings. Given the specialized nature of AI-related crises, the results might not translate seamlessly to crises in other domains or industries. Additionally, potential cultural biases may exist, as the participant group could predominantly represent specific cultural backgrounds. Consequently, the findings might exhibit variations if conducted in different cultural contexts. Additionally, the small number of participants, raises concerns about the generalizability of the results. A larger and more diverse sample would have allowed for a more robust statistical analysis and increased the external validity of the findings.

"Editing assistance provided by ChatGPT"

#### **List of References**

- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete Problems in AI Safety* (arXiv:1606.06565). arXiv. http://arxiv.org/abs/1606.06565
- Bowen, S. A., & Coombs, W. T. (2020). Ethics in crisis communication. In F. Frandsen & W. Johansen (Eds.), *Crisis Communication* (pp. 543–562). De Gruyter. https://doi.org/10.1515/9783110554236-027
- Cao, W. (2022). Image Semantic Analysis in Visual Media Art Using Machine Learning and Neural Machine Translation. ACM Transactions on Asian and Low-Resource Language Information Processing, 3522576. https://doi.org/10.1145/3522576
- 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
- Coombs, W. T. (2019). *Ongoing Crisis Communication: Planning, Managing, and Responding*. SAGE Publications.
- Coombs, W. T., & Holladay, S. J. (2002). Helping Crisis Managers Protect Reputational Assets: Initial Tests of the Situational Crisis Communication Theory. *Management Communication Quarterly*, 16(2), 165–186. https://doi.org/10.1177/089331802237233
- Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018).
  AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
- Gilligan, C. (1993). In a Different Voice: Psychological Theory and Women's Development. Harvard University Press.
- Kusumah, H., Zahran, M. S., Cholied, P. R., Alkusna, M. S., & Hafidhi, N. A. (2022). Deep Learning on Facial Expression Detection: Artificial Neural Network Model Implementation. *CCIT Journal*, 16(1), 39–53. https://doi.org/10.33050/ccit.v16i1.2518
- Neff, G., & Nagy, P. (2016). Automation, Algorithms, and Politics | Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, 10(0), Article 0.
- Overton, H., Kim, J. K., Zhang, N., & Huang, S. (2021). Examining consumer attitudes toward CSR and CSA messages. *Public Relations Review*, 47(4), 102095. https://doi.org/10.1016/j.pubrev.2021.102095
- Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd Edition). http://www.pearson.com/content/one-dot-com/one-dot-com/us/en/highereducation/product.html
- Schoofs, L., Claeys, A.-S., De Waele, A., & Cauberghe, V. (2019). The role of empathy in crisis communication: Providing a deeper understanding of how organizational crises and crisis

communication affect reputation. *Public Relations Review*, 45(5), 101851. https://doi.org/10.1016/j.pubrev.2019.101851

- Seeger, M. W. (2006). Best Practices in Crisis Communication: An Expert Panel Process. Journal of Applied Communication Research, 34(3), 232–244. https://doi.org/10.1080/00909880600769944
- Simola, S. (2003). Ethics of Justice and Care in Corporate Crisis Management. *Journal of Business Ethics*, 46(4), 351–361. https://doi.org/10.1023/A:1025607928196
- Tao, W., & Kim, S. (2017). Application of two under-researched typologies in crisis communication: Ethics of justice vs. care and public relations vs. legal strategies. *Public Relations Review*, 43(4), 690–699. https://doi.org/10.1016/j.pubrev.2017.06.003
- Ulmer, R., Sellnow, T., & Seeger, M. (2022). Effective Crisis Communication: Moving From Crisis to Opportunity. https://us.sagepub.com/en-us/nam/effective-crisiscommunication/book276847
- Walsh, G., & Beatty, S. E. (2007). Customer-based corporate reputation of a service firm: Scale development and validation. *Journal of the Academy of Marketing Science*, 35(1), 127– 143. https://doi.org/10.1007/s11747-007-0015-7
- Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing*, 60(2), 31–46. https://doi.org/10.2307/1251929

Appendix A. Study 1 – Stimuli		
Deny Strategy		
Title: Hexxa Denies Use of AI Recruiting Tool in Hiring Process		

A report emerged claiming that Hexxa's AI recruiting tool was found to have a gender bias in favor of male candidates. In response to the allegations, Hexxa issued the following statement:

"Hexxa strongly denies using the AI recruiting tool to evaluate any job applicants. We want to assure our customers, partners, and the public that our hiring process is fair and unbiased. The AI system in question was developed for research purposes only and has never been utilized by our recruiters in the decision-making process."

Excuse Strategy	
Title: Hexxa Addresses AI Gender Bias Controversy	

A report emerged claiming that Hexxa's AI recruiting tool was found to have a gender bias in favor of male candidates. In response to the allegations, Hexxa issued the following statement:

"Hexxa acknowledges the findings of the report but would like to provide some context regarding the situation. While it is true that the AI system may have exhibited a gender bias, we believe this to be an unintended consequence of the algorithm's design, rather than a reflection of our company's values or hiring practices. The development of the AI system was influenced by external data sources that inadvertently introduced bias into the system. We believe AI mistake on a gender bias might happened from time to time."

Apology Strategy

Title: Hexxa Acknowledges AI Gender Bias and Commits to Change

A report emerged claiming that Hexxa's AI recruiting tool was found to have a gender bias in favor of male candidates. In response to the allegations, Hexxa issued the following statement:

"We **sincerely apologize** for the shortcomings of our AI recruiting tool, and we acknowledge the concerns raised by the public. We recognize the importance of addressing this issue and are taking immediate steps to remedy it. We have halted the use of the AI system and will engage external experts to help us identify and correct any biases in its algorithms. Furthermore, we will invest in additional training for our recruiters to ensure a fair and unbiased hiring process. Hexxa is dedicated to continuous improvement and will learn from this experience to better serve our employees, customers, and partners."

Appendix B. Study 2 – I	Ethic of Care Message
-------------------------	-----------------------

Ethic of Care	(High)
---------------	--------

We understand the feelings of our publics and remain sensitive to the potential harm that such issues can cause. At Hexxa, we are committed to fostering a diverse and inclusive work environment, and we understand the importance of nurturing relationships among our publics. We will be sensitive and responsive to the feelings of those affected by this issue and consider the contextual complexities of our relationships with various stakeholders.

<u>As part of our action plan</u>, we will allocate an additional \$2 million to fulfill our ethical responsibilities and strengthen our commitment to diversity and inclusion. Furthermore, we are excited to announce a collaboration with the University of Pennsylvania to conduct indepth research on ethical responsibilities in the field of AI and recruitment.

Our focus will be on fulfilling our conflicting responsibilities and doing what is right for the victims, rather than solely concentrating on legalities. We pledge to keep our stakeholders informed of our progress and remain accountable throughout this process."

Ethic of Care (Low)

We understand the feelings of our publics and remain sensitive to the potential harm that such issues can cause. At Hexxa, we are committed to fostering a diverse and inclusive work environment, and we understand the importance of nurturing relationships among our publics. We will be sensitive and responsive to the feelings of those affected by this issue and consider the contextual complexities of our relationships with various stakeholders.

Our focus will be on fulfilling our conflicting responsibilities and doing what is right for the victims, rather than solely concentrating on legalities. We pledge to keep our stakeholders informed of our progress and remain accountable throughout this process."

Appendix C. Study 2 Stimuli Example- Apology Strategy with Ethic of Care (High)

Title: Hexxa Acknowledges AI Gender Bias and Commits to Change with Care

A report emerged claiming that Hexxa's AI recruiting tool was found to have a gender bias in favor of male candidates. In response to the allegations, Hexxa issued the following statement:

"We **sincerely apologize** for the shortcomings of our AI recruiting tool and acknowledge the concerns raised by the public. We will halt the use of the AI system and engage external experts to help us identify and correct any biases in its algorithms.

We understand the feelings of our publics and remain sensitive to the potential harm that such issues can cause. At Hexxa, we are committed to fostering a diverse and inclusive work environment, and we understand the importance of nurturing relationships among our publics. We will be sensitive and responsive to the feelings of those affected by this issue and consider the contextual complexities of our relationships with various stakeholders.

As part of our action plan, we will allocate an additional \$2 million to fulfill our ethical responsibilities and strengthen our commitment to diversity and inclusion. Furthermore, we are excited to announce a collaboration with the University of Pennsylvania to conduct in-depth research on ethical responsibilities in the field of AI and recruitment.

Our focus will be on fulfilling our conflicting responsibilities and doing what is right for the victims, rather than solely concentrating on legalities. We pledge to keep our stakeholders informed of our progress and remain accountable throughout this process."

### Appendix D. Measurement of Ethic of Care

I believe Hexxa ....

- 1) Commits to fostering a diverse and inclusive work environment
- 2) Fulfills ethical responsibilities to reflect publics' feelings
- 3) Emphasizes relationship nurturing among publics
- 4) Uses action plan to prevent conflicts to ensure fairness in future
- 5) Does what is right for victims rather than focusing on legalities
- 6) Remains sensitive to the potential harm that conflict can cause



## Appendix E. Impact of Response Strategies and Ethic of Care on Organizational Reputation

Appendix F. Impact of Response Strategies and Ethic of Care on Supportive Behavioral Intention

