

University of South Dakota

USD RED

Dissertations and Theses

Theses, Dissertations, and Student Projects

2021

**THE EFFECTS OF ARTIFICIAL INTELLIGENCE ON CRISIS
COMMUNICATION REBUILD STRATEGIES**

Rachel Ehlers

Follow this and additional works at: <https://red.library.usd.edu/diss-thesis>

**THE EFFECTS OF ARTIFICIAL INTELLIGENCE ON CRISIS COMMUNICATION
REBUILD STRATEGIES**

By

Rachel R. Ehlers

B.A., University of South Dakota, 2020

A Thesis Submitted in Partial Fulfilment of
the Requirements for the Degree of
Master of Arts

Department of Media & Journalism

Mass Communication Program
In the Graduate School
The University of South Dakota
December 2021

The members of the Committee appointed to examine
the Thesis of Rachel Ehlers
find it satisfactory and recommend that it be accepted.

DocuSigned by:
Travis Loof
61672B4333F94FE...

Chairperson

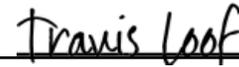
DocuSigned by:
Lori Costello
A9AEA2FFEA1445F...

DocuSigned by:
Elizabeth Manser Payne
BDFB732BFC9B4E1...

ABSTRACT

Organizational crises can hinder companies beyond repair if the repercussions of the crisis are not offset with a communications strategy. The severity of the crisis may vary, however, the response strategy used by an organization is a crucial factor in deciding how damaging the situation will be for the business. The types of responses, as described in Situational Crisis Communication Theory (Coombs, 2007), have been studied in conjunction with both the responses' crisis types and the communicator delivering the message. However, studies on the source of the crisis response have rarely gone past investigating the traditional response personnel. The current study explores the effectiveness of a nontraditional message source, artificial intelligence (AI), when delivering a rebuilding crisis response strategy. AI, while becoming more accepted into many facets, it is still questioned by many when used in situations with high stakes. This study investigates the reaction of the public if an organization used AI to respond to a major crisis. The current study could potentially influence future research in AI trust building and crisis communication response strategies.

Thesis Advisor

A handwritten signature in black ink that reads "Travis Loof". The signature is written in a cursive style and is positioned above a horizontal line.

Dr. Travis Loof

Table of Contents

Committee Signature	i
ABSTRACT.....	ii
Table of Contents	iii
List of Tables	vi
List of Figures	vii
Chapter 1: Introduction	1
Chapter 2: Literature Review	2
Situational Crisis Communication Theory.....	2
Background.....	2
Relationship and Crisis History	6
Rebuild Strategies & Their Effectiveness.....	8
Artificial intelligence	12
Background.....	12
Explainable Artificial Intelligence.....	12
AI and Human Computer Communication.....	14
Trustworthiness of AI.....	15
AI in Crisis Communication.....	20
Chapter 3: Method	21
Procedure	21

Participants.....	22
Stimuli.....	22
Measures	23
Trust in Message Source.....	23
Trust in Organization	23
Organizational Reputation	24
Source Credibility	24
Manipulation Check.....	24
Chapter 4: Results	25
Chapter 5: Discussion	27
Theoretical Implications	28
Practical Implications.....	29
Future Directions	30
Limitations	31
References.....	32
Appendix A.....	40
Appendix B.....	41
Appendix C.....	42
Appendix D.....	43
Appendix E	43

Appendix F.....	45
Appendix G.....	46
Appendix H.....	47

List of Tables

Table A1: *Message Source Trust t-Test*

Table B1: *Organization Trust t-Test*

Table C1: *Organizational Reputation t-Test*

Table D1: *Credibility t-Test*

List of Figures

Figure 1: *SCCT Crisis Response Strategies*

Figure 2: *Scope of Explainable Artificial Intelligence*

Chapter 1: Introduction

The study of communication dates back centuries. While many communication disciplines have been presented in literature for some time, crisis communication is a newer area comparatively. With crises being both sudden and unexpected, they pose a large threat to an organization's financial and reputational standing (Coombs, 2007). Even with the threat that crises pose to an organization, the study of crisis communication is dominated by case studies, telling us little about how an organization's stakeholders respond to crisis and crisis management strategies (Dean, 2004). More recently, however, academics have begun to understand how organization reputations are formed and affected by a crisis, and the ethical responsibilities that go with a crisis response from an organization (Coombs, 2007). As scholars continue to explore the best possible avenues for identifying and responding to a crisis, technology has become an essential tool for many organizations throughout these processes.

Artificial intelligence (AI) has rapidly advanced in the last 60 years (Nah et al., 2020). AI, while present in many areas of research, is still lacking in investigating communication studies concerning social, political, and cultural areas (Nah et al., 2020). Using AI in the crisis communication field, specifically within the domain of Situational Crisis Communication Theory (Coombs, 2007), has little literature to explore the various possibilities when combining the two subjects. With the fast-paced atmosphere of both communications and AI, it is likely that the two will become increasingly intertwined. It is important for scholars and professionals to know how AI can enhance or hinder strategies used within crisis communication. This study investigated if using a rebuild strategy from SCCT is still effective when the crisis responses are knowingly generated by an artificial intelligence source or a human source. Rebuild strategies are a communication method that organizations can use to gain reputational assets (Coombs,

2007). When using a human media representative to engage with stakeholders, rebuild strategies have been repeatedly studied and consistently been found as the most effective crisis strategy for maintaining the reputation of an organization when presented with a preventable crisis. Within this study, the crisis response will be coming from either an AI media representative or a human representative and studied accordingly. Traditional crisis communication calls for an employee of high standing in an organization to deliver the response within a crisis (Vercic et al., 2018).

Chapter 2: Literature Review

Situational Crisis Communication Theory

Background

Situational crisis communication theory (SCCT) was created to understand the public's perception and approval of an organization following a crisis or paracrisis (Sellnow & Seeger, 2013). This theory supplies an evidence-based framework for organizations to use when trying to protect their reputation throughout and following a crisis (Coombs, 2007). SCCT was first developed by Timothy Coombs in 1995. Coombs wanted to enhance the topic of crisis within public relations research. Before the development of SCCT, the exploration of crises within an organization was limited to case studies of the past with no real framework for understanding how to protect the organization and their reputations during a crisis. SCCT's framework is based on Attribution Theory (Coombs, 2004; Weiner, 1985). Attribution theory looks at the way that people search for the cause of events, more so when the events are unexpected or negative, which is the case with many crisis events (Coombs, 2007).

Coombs presents three factors that influence stakeholders' attributions: initial crisis responsibility, crisis history, and prior relational reputations. Initial crisis responsibility is the attributions of the stakeholders regarding the organization's personal control for the crisis- how

much the actions of the organization caused the crisis. The initial assessment of crisis responsibility is based on how the crisis is being framed (Coombs, 2007). Research has shown that increased attributions of crisis responsibility by stakeholders leads to lower reputational scores for the organization (Coombs & Holladay, 1996). Crisis history is based on whether the organization has faced similar crises in the past (Coombs, 2007). Past crises can be looked at as a potential indicator for a pattern of undesirable behavior. The larger the crisis history of an organization, the more crisis responsibility is produced and the greater reputational threat (Coombs, 2004). Research has shown that there is a weak correlation between crisis history and the perceptions of crisis responsibility but a strong correlation between crisis history and organizational reputation (Coombs, 2004). Prior relational reputation refers to how well an organization is perceived to have treated stakeholders prior to the crisis (Coombs, 2007). An unfavorable prior relational reputation, or a history of the organization treating stakeholders poorly, suggests little consideration from the organization for stakeholders across several domains, not just in the crisis itself (Coombs, 2007). Negative relationships with stakeholders have been believed to intensify attributions of crisis responsibility. The prior relational reputation also has been shown to have an indirect effect on the reputational threat (Coombs, 2004).

Based on these attributions, the theory then describes three crisis clusters: victim, accidental, intentional. Within the victim cluster, the organization is a victim of the crisis along with the stakeholder. This cluster often has very weak attributions of crisis responsibility (Coombs & Holladay, 2002). Crises that would fall into the victim cluster would include natural disasters, rumors, workplace violence and product tampering (Coombs, 2007). The accidental cluster represents crises that arise from unintentional actions by the organization (Coombs & Holladay, 2002). Crises in this cluster would include technical-error accidents, technical-error

product harm and challenge. This cluster has minimal attributions of crisis responsibility as the event is considered unintentional or uncontrollable on the organization's behalf (Coombs, 2007). Finally, the intentional, or preventable, cluster refers to crisis in which the organization either purposefully places stakeholders at risk, knowingly took inappropriate actions or human error that might have or could have been avoided (Coombs & Holladay, 2002). Crises that fall into this cluster include human-error accidents, human-error product harm and organizational misdeed. The intentional cluster has very strong attributions of crisis responsibility (Coombs, 2007). In the case of this study, a preventable crisis cluster is used to provoke participants to place crisis responsibility with the organization.

Finally, there are three main strategies laid out by the theory for handling a crisis: denial, diminishment and rebuilding (Coombs, 2007). Each area is dependent on the crisis at hand, the organization, the stakeholders, and any other constraints surrounding the situation. The main objectives of crisis response strategies are (1) shape attributions of the crisis, (2) change perceptions of the organization in crisis and (3) reduce the negative affect generated by the crisis (Coombs, 1995). Deny strategies seek to establish a crisis frame and attempt to remove any connection between the organization and the crisis. If the organization is in fact not involved in the crisis, little damage will be suffered by the organization. Deny strategies are often used for victim cluster crises. Crises that happen as a result of a rumor or a challenge, crisis managers tend to argue that there is no 'real' crisis and must deny the truth to the rumor or refute any charges of immoral conduct (Coombs, 2007). Scapegoating, or blaming someone else for the crisis, also denied any organizational responsibility for the crisis (Coombs, 2006). If stakeholders choose to accept the no crisis, or scapegoat frame, little to no reputational harm will be done (Coombs, 2007). Diminish responses reflect attribution theory's contribution to SCCT. With this

strategy, crisis managers would accept that the crisis occurred and that the organization is involved but not to the extent stakeholders may originally believe (Coombs, 2006). When managers lessen an organization's connection to the crisis or have stakeholders view the crisis less negatively, the potentially harmful effects of a crisis are reduced (Coombs, 2007). Failure is possible if crisis managers cannot provide solid evidence to support their claims and/or cannot get stakeholders to accept the presented frame. Diminish strategies are often used when an accidental cluster crisis has occurred and typically uses excuses or lack of intent to reaffirm a crisis. While diminish strategies can be used with intentional crises, it will likely be much harder and more expensive to manage reinforcing a diminish frame (Coombs, 2007). If an organization needs to generate new reputational assets, rebuild strategies are often the best option. With a rebuild strategy, the goal is to offset the negatives from the crisis with the presentation of new positive information about the organization or reminding stakeholders of past good works (Coombs, 2007). Common ways of taking positive action to offset the negatives are offering compensation or a full apology on behalf of the organization. Rebuild strategies are used when the crisis at hand presents a severe reputational threat. Often this would include crises that fall into the intentional cluster or an accident cluster crisis that is paired with a crisis history or an unfavorable prior relationship reputation (Coombs, 2007).

Bolstering is a secondary strategy that is best used in addition to one the three primary strategies discussed (Coombs, 2007). Bolstering offers only minimal opportunity to develop reputational assets. Within the bolstering strategy, managers, who already have positive relations with stakeholders, call upon that goodwill to help protect the organization. This is often done by praising stakeholders for their efforts during the crisis as a way to improve the relationship between organization and stakeholder, or the organization plays the victim role in the crisis to try

and invoke sympathy with the stakeholders. An organization may also use past good works to offset the negatives that arise from the crisis (Coombs, 2007). Image restoration theory (Benoit, 1995), another framework theory of SCCT, claims that bolstering is most effective when an organization is accused of unjust actions and wants to minimize the offensiveness of those actions. An example of bolstering being used is in a response from an Exxon chairman after the Valdez oil spill: 'Exxon has moved swiftly and competently to minimize the effect this oil will have on the environment, fish and other wild-life' (Benoit, 1997, p. 180). In this situation, Exxon was using the good deed of minimizing harm to the environment to offset the fact that the organization itself was the reason for said harm.

Relationship and Crisis History

There are three factors within SCCT for influencing stakeholder's attributions: initial crisis responsibility, crisis history, and prior relational reputations. Crisis history indicates whether the incident was a one-time event or part of a pattern (Coombs, 2001). Coombs and Holladay (2001) investigated how the relationship history and the performance history fuses together. Here the authors state "relationships affect the crisis by shaping perceptions of the crisis and the organization in the crisis" (Coombs & Holladay, 2001, p. 321). Coombs and Holladay (2001) provided the model of the variables they studied on relationship history and crisis history and how they interact with each other. The results of this study showed that relationship history does play a role in how stakeholders perceive a crisis and an organization. An unfavorable relationship history leads people to place more crisis responsibility on the organization. Relationship history accounted for three percent of variance in crisis responsibility and crisis history represented four percent (Coombs & Holladay, 2001). The authors found one main difference between crisis and relationship history. While both contributed similarly to crisis

responsibility, relationship history factored more heavily into organizational reputation than crisis history (Coombs & Holladay, 2001). Coombs also examined crisis history and how past crises act as an intensifier for crisis responsibility and the reputational threat across the entire range of clusters (Coombs, 2004). The study found crisis history, as a form of consistency, indicated how the organization may repeat a certain behavior. This specific study looked at participants who were considered non-victims. Non-victims are described as those who were not injured by a specific crisis by following the crisis through the news media and may have weaker held perceptions of the organization in question, as compared to victims who experienced the crisis firsthand within the organization (Coombs, 2004). Non-victims were also used in this current study. The crisis used for this experiment spanned across all the crisis clusters and was manipulated to have one of three specific characteristics: a history of a similar crisis, no presentation of information about past crisis, or information indicating no past crisis. Coombs (2004) states that having the three different scenarios, including the neutral one was important because many times in a crisis, stakeholders find themselves knowing very few specifics of an organization concerning a current crisis. It was also important for Coombs to be able to determine the extent to which crisis history and information presenting past crises would produce effects that would differ from results produced by the neutral condition. Coombs looked into two clusters of crisis, victim, and accident. Within these clusters, he chose two crises from each cluster. Workplace violence and product tampering were selected from the victim cluster. Rumor and natural disaster were not selected from the victim cluster because rumors create a unique dynamic with the need to deny the crisis and natural disasters have been previously found to not be affected by a severity intensifier (Coombs & Holladay, 2004). Technical accident and technical-error recall were selected from the accident cluster. Challenge was not selected from

the accident cluster for the same reasons as rumor. The results of this study found that product tampering and technical-error product recalls with the condition of similar crisis history led to strong attributions of crisis responsibility and low perceptions of the organization's reputation. Product tampering paired with crisis history was used in the current study stimuli to emulate the threat to the organization's reputation. With a large reputational threat present, organizations would need to be overly strategic within their response strategies to attempt to offset the crisis responsibility and current damage to their reputation. There was no major difference when this type of crisis was tied to the unknown condition or the no crisis history condition. In a follow-up analysis, it was also found that workplace violence and technical-error accidents had the same results as the previous crises, organizations with a similar crisis history were held more responsible thus would need an effect response to pull the organization out of the crisis effectively.

The idea of crisis and relationship history playing a role in the attributions of stakeholders is an idea that was put into use before Coombs called it crisis history and implemented it into SCCT. Griffin, Babin, and Attaway (1991) investigated the impact of negative publicity on the attitudes and intentions of stakeholders. Within this study, they described the preconceived idea of performance history. These authors state that when stakeholders are exposed to a situation multiple times, one can gain more information to apply to the attribution process. When multiple occurrences of a specific problem happen, the organization is more likely to be perceived as the cause of the problem (Griffin et al., 1991). These assumptions correlate with those that were later depicted in SCCT and further studies.

Rebuild Strategies & Their Effectiveness

According to Coombs (2007), rebuild strategies are the primary way for organizations to generate new reputational assets. Within this strategy, crisis managers will say and do things that will directly benefit stakeholders in hopes that positive actions will offset the crisis. Many times, this strategy takes the form of offering compensation or a full apology. Rebuild strategies are used when a crisis poses a severe reputation threat. A severe reputational threat would include any instance that is seen as extremely offensive and/or contradicts a key factor that the organization has used to build its reputation (Coombs, 2002). For example, if an organization used concern for the environment as a key factor in building their reputation, the revelation of illegal dumping of toxic chemicals would cause stakeholders to see the organization as hypocritical thus shattering their reputation. Crises that could inflict this level of reputational damage could be an intentional crisis, or an accidental crisis paired with a crisis history.

Claeys, Cauberghe, and Vyncke (2010) used SCCT to study how different responses influence people's internal and external locus of control. Locus of control refers to the attributions that people make for behavioral consequences and reinforcements (Duffy et. al., 1977). Figure 1 shows Coombs (2007) matches between crisis types and strategies.

Figure 1

SCCT Crisis Response Strategies

Crisis types	Crisis response strategies
<i>Victim cluster</i>	<i>Deny strategies</i>
Natural disaster	Attack the accuser
Rumor	Denial
Workplace violence	Scapegoat
Product tampering/Malevolence	
<i>Accidental cluster</i>	<i>Diminish strategies</i>
Challenges	Excuse
Technical-error accidents	Justification
Technical-error product harm	
<i>Preventable cluster</i>	<i>Rebuild strategies</i>
Human-error accidents	Compensation
Human-error product harm	Apology
Organizational misdeed with no injuries	
Organizational misdeed management misconduct	
Organizational misdeed with injuries	

This study specifically investigated the effectiveness of the rebuild strategy. These scholars choose one crisis from each of the three clusters and choose two strategies matched to each cluster. This study found that the reputation of an organization using rebuild strategies is more positive than using a deny or diminish strategy (Claeys et al., 2010). The same result was found by Susco (2012) when looking at nonprofits in crisis. Participants of this study evaluated the image of the nonprofit more positively and were more likely to participate in the organization when a rebuild strategy was used in a preventable and accident crisis.

Additionally, Vercic, Vercic, and Coombs (2018) expanded the literature on response strategy by incorporating differing communication sources. This research focused on how communicative responses and various sources of communication affected stakeholders' perceptions of the organization. Participants were presented with four distinct types of response situations: the apology of a CEO, apology of an official spokesperson, the sympathy of a CEO, and sympathy of an official spokesperson (Vercic et al., 2018). The students were asked to assess three different organizations on the cover page of the study's questionnaire to assess the organization's reputation prior to the participants being exposed to the crisis. After filling out the questionnaire, participants watched a news video clip that resembled a typical news item but involved a fictitious company that was involved in an accident where one of its oil tankers caused extreme pollution near the coast. The video report included a crisis account from a

journalist followed by an explanation by either the spokesperson or CEO. In this study, the role of the spokesperson and CEO were played by the same actor just differing titles depending on the scenario. In two of the cases, the actor representing the CEO/spokesperson apologized and took full responsibility. In the other two, he stated sympathy about what happened but without taking any responsibility. After viewing the clip, participants then filled in the rest of the questionnaire. Out of the four conditions mentioned previously, when the official spokesperson apologized there was no change in the organization's reputation. Both situations that used sympathy, both from the CEO and the spokesperson, showed a slight decrease in reputation. There was no significant difference between the CEO apology and the spokesperson apology. Overall, the results also did not show a difference between deciding between a CEO and a spokesperson to deliver a crisis strategy. However, apology was more effective than sympathy throughout all the conditions. An apology strategy would fall into the rebuild strategy when using SCCT. In this case, the rebuild strategy led to more positive impressions, thus would help the reputation of the organization post-crisis.

This study was significant because past research (Pauly & Hutchison, 2005; Seeger et al., 2003) placed heavy importance on having someone in power, such as a CEO, involved in the crisis. It was argued that having an individual of power as the face of the organization in a crisis increased stakeholders' confidence thus increased the organization's chance of coming out of the crisis with minimal reputational damage. However, within this study, there was no significant difference. The authors describe the notable difference between a sympathy approach from a CEO vs. a spokesperson. The results indicated that when stakeholders perceived the CEO as not taking accountability for the crisis within their organization, stakeholders wanted to see blame

fall solely to the CEO. Respondents did not hold the spokesperson to the same level of responsibility as the CEO when placing blame for the crisis (Vercic et. al., 2018).

Artificial intelligence

Background

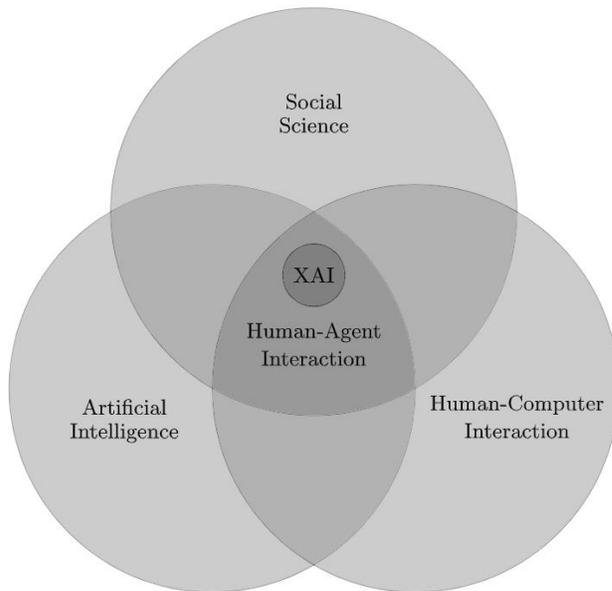
Artificial intelligence (AI) is an area of computer science where programmers attempt to create intelligent machines that can behave, think, and make decisions like humans. Artificial intelligence is best described as a “thinking power made by humans” (Dhankar & Walia, 2020). This area of science was formed with the idea that one day machines will be able to replicate human characteristics. AI has several stages consisting of planning, reasoning, analyzing data, predicting outcomes, and acting accordingly (Dhankar & Walia, 2020). There are also some basic elements that makeup AI. These elements are the human element, the knowledge base, and the algorithm set (Dhankar & Walia, 2020). Being able to imitate human intelligence requires a focus on several qualities. These qualities work together in different facets versus being separate. These components consist of learning, problem solving, logic and reasoning, perception, language understanding, neural network, and machine learning (Dhankar & Walia, 2020).

Explainable Artificial Intelligence

Explainable artificial intelligence (XAI) has become increasingly popular in the social sciences in the past decades. The idea behind XAI is that if AI is more transparent, interpretable, or explainable, users will be able to understand AI better and in turn, trust AI more (Miller, 2019). Miller (2019) decides a few ways to increase the transparency between intelligent agents and others. Two of these ways include generating decisions while considering how well humans will be able to understand the decision and explicitly explaining decisions to people. Figure 2 shows the scope of XAI.

Figure 2

Scope of Explainable Artificial Intelligence



Miller's studies focus on the social science circle of the XAI diagram and how philosophy, social and cognitive psychology, and cognitive science views artificial intelligence explanations (Miller, 2019). Miller, (2019) found four major findings believed to be important to XAI. They concluded that explanations are contrastive, explanations are selected, probabilities do not matter, and explanations are social. When explanations are contrastive, they are sought in response to counterfactual cases. This means that people do not ask why an event happened but rather why it happened instead of another event. Selected explanations are often done so in a biased manner. People rarely expect an explanation to consist of an actual and complete cause of an event. Rather, people will select a number of causes to be the explanation. It was also found that probabilities likely do not matter. Referring to probabilities or statistical relationships in an explanation are not as effective as referring to causes. Finally, explanations are social. These explanations transfer knowledge by presenting them as part of a conversation or interaction. This allows for AI to be presented relative to the explainer's and explainee's beliefs (Miller, 2019).

Miller (2019) continues to explain that explanations are not just causal attributions but are contextual. People tend to only care about what parts of the event (AI) are relevant to the person seeking the explanation (Miller, 2019).

AI and Human Computer Communication

As curiosity in AI grows, so do the studies about how human-AI communication and human-human communication differ. McDowell and Gunkel (2016) posed the predominant question “what can, or should we do in the face of this other—this other form of otherness that calls into question everything we thought we knew about the communicative subject and the subject of communication” (p.2).

Within the field of human and computer communication, there are two main categories of visual representations of people, avatars, and agents, described by Fox et al. (2014). Avatars are controlled by humans while agents are controlled by computer algorithms (p. 402-403). The difference between the two representations lies in the element of who is in control. This control also gives avatars and agents individual characteristics. Agents do not require the presence of a human being to be successful. Once an agent is built, it can be replicated easily and does not need the benefits that a human-controlled avatar would need such as sleep, time off and a paycheck. However, it has been theorized that agents are not as persuasive as avatars can be (Blascovich, 2002). Much of the effectiveness attributed to agents or avatars depends on how well they are perceived by users. Fox et al (2014) found that there was a significant overall effect of avatars being more influential than agents (p. 413). Prior study results also indicated that regardless of any actual difference in performance, if users believed they were interacting with a human instead of a machine, the users became more susceptible to social influence (Lim & Reeves, 2010; Okita et al, 2008). A practical way to increase social influence in an agent is for

designers to give agents more markers of humanness. This would include features that can be processed and categorized as human such as natural speech patterns, emotions and flexibility. It would be crucial for designers to avoid highly scripted behaviors and add personalizing and tailoring details into the interaction (Fox et al, 2014). It is also important to understand the benefits of a hybrid system. Cognitive agents can play a vital role in balancing the impact of a human's cognitive capacity. This was especially true for decision-making teams under various time-stressed situations (Fan et al, 2010).

Further, Mou and Xu (2017) studied the comparison of the initial human-AI social interaction with the initial human-human social interaction. Specifically, they compared personality traits and communication attributes reflected in the two interactions. The study indicated that the participants demonstrated different personality traits when communicating with AI than with humans. Researchers used a chatbot designed for a 17-year-old girl with a lively and outgoing personality and all the communication took place over an online chat platform. The participants were found to be more open, agreeable, extroverted, and conscientious when interacting with another human over AI. The only area in which no significant difference was detected was the level of control felt over the social interaction (Mou & Xu, 2017). This study concluded that users did in fact apply different communication strategies when interacting with AI over humans.

Trustworthiness of AI

AI has become increasingly present in many decision-making processes historically taken on by humans. These models are being used to help humans come to a fair and accurate judgment in many instances such as college admissions, prison sentences, hiring, and mortgage approvals (Ashoori & Weisz, 2019). Since AI is starting to play a hand in these life changing

decisions, it is important to understand how trustworthy people perceive this process to be. Ashoori and Weisz (2019) examined seven factors that defined the trust boundaries of AI-infused decision making. The boundaries include stakes, decider, trainer, model interpretability, train and test set description, social transparency and model confidence. The stakes boundary refers to what consequences would accompany the decision at hand. Research has considered both low-stakes decisions and high-stakes decisions, especially as AI is being increasingly used in high-stake situations (Murawski, 2019; Rudin, 2018). The decider boundary decides who is responsible for making the ultimate decision. Decisions made entirely by AI are compared to decisions made by human supported AI (Ashoori & Weisz, 2019). The trainer boundary discusses how AI was trained. Automated training methods have become increasingly popular (Feurer et al., 2015) thus researchers have considered AI models trained by human data scientists and ones trained by automated AI. Some kinds of AI models are interpretable using things like decision trees or rule-based scoring systems. For these interpretable models, the process by which the model arrived at a recommendation can be examined. Others are deep neural networks, known as black boxes, whose inner workings do not give insight into how the model came to a recommendation. Although it is recognized that explanations can be generated for black box models, it is considered more desirable if the model is interpretable due to criticisms of black box models (Rudin, 2018). The train and test set description boundary refers to whether or not a model comes with a description of how it was trained and tested. Advocates for AI fairness have called for transparency into how an AI model is trained in order for others to evaluate the trustworthiness (Bellamy, 2018). Social transparency refers to seeing how models make decisions for others. Having this transparency may be an indicator of whether the model exhibits bias (Ashoori & Weisz, 2019). The final boundary discussed, model confidence, alludes

to whether the confidence a model has in its decision is visible to others. Many types of AI models provide some estimate of their confidence levels (Ashoori & Weisz, 2019). The researchers also evaluated trust as a multi-dimensional concept looking into overall trustworthiness, reliability, technical competence, understandability and personal attachment. Researchers thought that the process ought to be trusted and studied the overall perceived trustworthiness of the decision-making process. The process results also expected consistent outcomes thus reliability was studied (Madsen & Gregor, 2000). Technical competence was studied to make sure AI was being used appropriately and correctly (Madsen & Gregor, 2000). Understandability made sure that participants understood how the process worked (Madsen & Gregor, 2000) and personal attachment studied how the participants liked the study (Hoffman et al., 2018).

In the overall ratings of trust, the four factors of trust fell in the middle of the scale with a standard deviation of 2.2 indicating a possibility of skepticism with use of AI scenarios. The stakes boundary exhibited a large effect on trust. Generally, the higher-stakes scenarios generated a lower trust of AI. Model interpretability and train and test set description also had strong effects on each of the trust factors. Interpretable models were preferred over black box models on a constant basis. Social transparency saw a smaller effect on the four trust facets than the boundaries above but was still shown to impact trust. Within the decider boundary participants expressed a mix of opinions towards a human or an AI making the ultimate decision in the scenarios given. Some participants felt comfortable with AI being the ultimate vote if the process to reach a decision was transparent. Others expressed concerns regarding AI and a final decision. Participants also generally preferred that the AI model in use was trained by a data scientist rather than an automated method, however, the effect size differed. This small

significance was also seen in model confidence. The reported model confidence was higher when AI was responsible for the final decision. The confidence lowered when a human was factored in. Regardless, participants felt strongly about having information regarding the model's confidence (Ashoori & Weisz, 2019).

Jakesch et al. (2019) examined how AI involvement affects trust in an online profile context. This study investigated the impact of trust on online self-presentations when AI was suspected of being involved with the creation of the profile. In this case, AI-Mediated Communication (AI-MC) was explored. AI-MC refers to interpersonal communication that is augmented by an algorithm to achieve a specific communication or relationship goal and can heavily affect interactions ranging from messaging one-to-one to one-to-many (Jakesch et al., 2019). Researchers used Airbnb, a website where hosts can list properties that guests can book and rent to present guests with profiles that were suspected to be either entirely human written, entirely AI-generated or a combination of the two. The trustworthiness of the hosts' profile was then scored to see the impact of using AI on this platform. Study one presented either all human-written or all AI-generated profiles. This study found that there were no significant differences in the perceived trustworthiness of the two profiles despite the belief of all human or AI involvement. Study two was conducted using the same set up as study one but instead researchers manipulated the ability for participants to know who wrote the profile. In the profiles that were believed to be AI-generated, guests scored the hosts as less trustworthy than profiles believed to be written by humans (Jakesch et al., 2019). This indicated that AI-MC does in fact influence perceived trustworthiness. When the guests were told that all profiles were written by AI, the guests were willing to trust the profile to the same degree that they did when told the profiles were all human written. The transparency of how the profile was generated lead to a

higher level of trust for the host. On the contrary, when guests were exposed to a mix of profiles, the profiles suspected of being AI-generated were considered less trustworthy. According to Jakesch et al. (2019), this phenomenon has been labeled the Replicant Effect. The Replicant Effect says that in a mixed source world, the knowledge, or even suspicion that a profile is AI-generated leads to distrust. Knowing that trust in AI varies in online profiles, it can then be called into question whether this trust would translate into a crisis situation if the same level of transparency is given. Thus, it poses that:

H1: Individuals are less likely to trust AI as an information source than a human

H2: Individuals are less trusting of an organization when AI was used as a representative than a human representative

H3: The perceived reputation of an organization will decline when AI is used as a representative

Twitter is known for its automated agents, for example, Twitter provides chatbots with an API key to interact with the platform (Edwards et al., 2014). The automation of a Twitter feed often blurs the lines between a human user and a bot user (Chu et al., 2010). Edwards et al (2014) conceptualized this behavior as the term “botification.” Additionally, Edwards et al. (2014), examined whether a known Twitter-bot agent is perceived differently than a human agent on variables related to perceptions of communication quality. Communication quality was defined as source credibility (McCroskey & Teven, 1999), interpersonal attraction (McCroskey & McCain, 1974), computer-mediated communication competence (Spitzberg, 2006) and intent to interact (Edwards et al., 2014) were all studied to determine varying perceptions. Several studies have also examined Twitter in relation to communication quality, noting source credibility as one of the most important factors (Zhao, & Rossen, 2009; Westerman et al., 2012;

Morris et al., 2012). Significant differences were found between the treatment of the human and Twitterbot dependent variables of competence, character and caring dimensions of credibility, task and social interactions (Edwards et al., 2014). While the Twitterbot agent scored generally lower than the human agent, the bot's mean was still placed above the scale midpoint demonstrating that users perceived the Twitterbot as a credible source (Edwards et al., 2014). Participants rated the human agent higher for social and task attraction. Finally, this study found no significant difference in perceptions of credibility, communication competence and intent to interact between the human agent and Twitterbot agent (Edwards et al., 2014). This could suggest that AI as a source of information could still be deemed credible by users. If users are deeming AI sources within social media as credible, it could be questioned whether AI would also be seen as credible within a crisis response. Thus, is it posed:

H4: Individuals will perceive the overall credibility of an AI representative lower than a human representative

AI in Crisis Communication

Communications and artificial intelligence are more closely related than many give credit for. According to Gunkel (2017), it is communication that supplies the science, with its definitive test case and experimental evidence, in order to blend both communication and AI together. Lee, Nah, Chung, and Kim (2020) explored the perceived credibility of AI news media. Media credibility deals with how believable and trustworthy stakeholders view communication. In this study, a survey was distributed that specifically asked how much respondents consumed different types of media and their perceptions of AI news credibility. The scholars found a positive effect on AI news credibility in all media outlets except for newspapers. Specifically, online news sites had the highest perception of credibility. Lee et al. (2020) credit this to the fact that online news

sources produce a higher number of algorithm-based news. Following online sources, AI was perceived most trustworthy in TV news. This study allows for testable theory for AI credibility within other types of media, or specific areas of media communication, such as crisis communication strategies.

Previous literature (Ashoori & Weisz, 2019; Bellamy, 2018; Jakesch et al., 2019) resulted in people having a mixed level of trust when AI becomes involved. This is especially true when people are not fully aware of the level of AI involvement (Jakesch et al., 2019). The level of trust a person places in an organization's content plays a significant role in developing an effective crisis strategy. As a crisis strategy is put into place, organizations must try to keep or even gain the trust of the public in order to navigate the situation successfully. Rebuild strategies have been found to be one of the most effective response strategies when dealing with a preventable crisis such as the one presented in this study (Claeys et al., 2010). The organization's crisis history is also included to solidify the need for a rebuild strategy to be used. Within a rebuild strategy, a relationship history is key (Coombs & Holladay, 2001). If the public does not trust the source of the message being delivered, the strategy becomes ineffective. As the world continues to become a mixed-source world (Jakesch et al., 2019), organizations will continue to consider integrating AI into their crisis response strategies.

Chapter 3: Method

Procedure

This study was approved by the researcher's IRB. Participants read the stimuli given at the beginning of the survey and confirmed their comprehension of the situation via a knowledge check (Berinsky et al., 2014). The knowledge check was "who created the press release in

response to the crisis.” Following the stimuli, participants were asked to complete the questionnaire, and then debriefed.

Participants

This study used a convenience sampling. 186 undergraduate students were recruited from a medium-sized Midwestern university. The respondents were offered extra credit for their participation in the study. Participants were between the ages of 18 and 29 years ($M = 19.6$, $SD = 1.87$). Participants were 42% male ($N = 73$), 57% female ($N = 100$) and >1% prefer not to disclose ($N = 2$). The majority of the participants identified as white ($N = 169$), followed by black or African American ($N = 5$), American Indian or Alaskan native ($N = 4$), Asian ($N = 4$), mixed race ($N = 2$) and self-reported ($N = 2$).

Stimuli

The stimuli were created using both a fictitious organization and crises to ensure that respondents of this study have not been exposed to the crisis situation or response prior to the study. Subjects were randomly assigned to one of two conditions.

The crisis being used for the stimulus was presented on the cover page of the survey. Vercic et al., (2018) used a similar manipulation. This situation consisted of an organization that has been accused of distributing contaminated products to the public (see Appendix E). Following more investigation, the organization found that product tampering had in fact taken place and as a result, 50 people have reported being hospitalized after ingesting the product. The stimuli included a mention of the organization’s crisis history to ensure that the organization's reputation was at stake. This crisis was presented in a consistent manner for all surveys. Following the explanation of the crisis, participants were instructed to read the company's response. The response was given in the form of a press release with either a media

representative Pat Kelly (see Appendix F) or ChatBot AI Agent (see Appendix G) as a contact person. Regardless of which media representative was used, the press release included the same wording, an admission of guilt, and a full apology as described in SSCT rebuild strategy. The identity of the media representative was placed in the response. Following the reading of the situation and strategy, participants were then directed to complete the rest of the survey.

Measures

This study used a modified measure adapted from various sources. A two-condition post-test experiment was used with message source (artificial intelligence and spokesperson) as the independent variable. Participants were then randomly assigned one of the two conditions: a press release including ChatBot AI Agent as a media representative and a press release including Pat Kelly as a media representative. The dependent variables were trust (in message source and in the organization), perceived reputation of the organization following the response and source credibility. All measures are located in Appendix H.

Trust in Message Source

To measure the participants' trust in the message source following the stimuli, questions originally used by Song et al. (2018) were measured on a 5-point Likert scale ($\alpha = .86$). Questions measuring trust included the following: "I trust the source to provide the best available information on the crisis", "I trust the source to provide enough information to allow me to decide my stance on the crisis", "I trust the source to provide truthful information about human safety regarding the crisis" and "I trust the source to provide timely information regarding the crisis."

Trust in Organization

To measure the participants' trust in the organization following the stimuli, six items developed by Hon and Grunig (1999) were measured on a 5-point Likert scale ($\alpha = .79$). Questions measured included the following: "This organization treats people like me fairly", "this organization can be relied on to keep its promises", "this organization has the ability to accomplish what it says it will do" and "I feel confident about this organization's skills" (Hon & Grunig, 1999). Participants will self-report their trust in the organization.

Organizational Reputation

To measure how participants perceived the reputation of the organization, questions were measured on a 5-point Likert scale ($\alpha = .2$) using five items from Coombs and Holladay's (2002) organizational reputation scale. The items present in the study included "the organization is concerned with the well-being of its publics", "the organization is basically dishonest", "I do not trust the organization to tell the truth about the incident", "under most circumstances I would be likely to believe what the organization says" and "the organization is not concerned with the well-being of its publics" (Coombs & Holladay, 2002) Participants will self-report their perception of the organization's reputation.

Source Credibility

To measure perceived credibility of the message source, a five-item credibility scale from Metzger, Flanagin, and Zwarun (2003) was used. This scale was measured on a 5-point Likert scale (none at all- a great deal) ($\alpha = .74$). This scale is based on a literature review performed by Flanagin and Metzger (2000) who identified believability, accuracy, trustworthiness, bias, and completeness of information as the most consistent measurements of source credibility.

Manipulation Check

To ensure that respondents understood the crisis and response, two screener questions were asked. One question was placed immediately following the stimuli, while the other was asked in the middle of the survey. This was done to gauge the comprehension of the stimuli throughout the study. Screeners work by instructing subjects to demonstrate that they understand the instructions and stimuli (Berinsky et al., 2014). The two screeners asked were who created the press release in response to the crisis and what product was tampered with in the crisis.

Chapter 4: Results

Participants who failed to correctly identify the author of the press release and the item that was tampered with were removed from analysis. In total, 81 responses were thrown out due to failing the screener questions. In addition, any participant that had not completed 97% or higher of the survey was also thrown out. 97% ensured that the majority of the questions, including both screener questions, were shown and answered by the participant. A Bonferroni correction was used to prevent any Type 1 errors due to the data being tested multiple times. Each test was designed to a $p=.05$ and four separate tests were run. The significant level following the correction was $p_{\text{new}}=.0125$.

Hypothesis one predicted participants would be less trusting of an AI information source in a crisis situation than a human information source. A one-tailed t-Test assuming unequal variances was run (see Table A1). There was not a significant difference in message source trust between participants who were exposed to the AI stimuli ($M= 3.22, SD=.84$) and participants who were exposed to the human stimuli ($M=3.48, SD=.81$); $t_{(180)} = -2.14, p = .02$. The results were compared to $p = .0125$ after data corrections. This means that participant's trust in AI as an information source was not statistically significant compared to their trust in the human message source thus rejecting **H1**.

Hypothesis two predicted participants would be less trusting of an organization when AI was used as a media representative than a human representative. A one-tailed t-Test assuming unequal variances was run after the data was cleaned (see Table B1). There was not a significant difference in organization trust between participants who were exposed to the AI stimuli ($M= 3.46$, $SD= 1.38$) and participants who were exposed to the human stimuli ($M= 3.39$, $SD=.76$); $t_{(133)} = .42$, $p = .34$. The results were compared to $p = .0125$ after data corrections. Participant's trust in the organization did not differ regardless of using AI or a human media representative. **H2** was rejected.

Hypothesis three predicted the perceived reputation of an organization would decline when AI is used as a representative. A one-tailed t-Test assuming unequal variances was run after the data was cleaned (see Table C1). There was not a significant difference in reputation between participants who were exposed to the AI stimuli ($M=3.24$, $SD=.55$) and participants who were exposed to the human stimuli ($M=3.24$, $SD=.46$); $t_{(172)} = .001$, $p = .5$. The results were compared to $p = .0125$ after data corrections. The overall perception of the organization's reputation did not differ based on the two media representatives that were used. **H3** was rejected.

Hypothesis four predicted participants would perceive the overall credibility of an AI representative lower than a human representative. After the data was cleaned, a one-tailed t-Test assuming unequal variances was run (see Table D1). There was not a significant difference in credibility between participants who were exposed to the AI stimuli ($M=2.94$, $SD=.71$) and participants who were exposed to the human stimuli ($M=2.92$, $SD=.67$); $t_{(179)} = -.22$, $p = .41$. The results were compared to $p = .0125$ after data corrections. Participants found both the AI and the human representative as a credible source in the crisis situation. **H4** was rejected.

Chapter 5: Discussion

As artificial intelligence continues to develop, it will continue to become more involved in all facets of communication, including crisis communication. Crisis communication places a large amount of trust in how well the organization's developed strategy will perform. Trust and perception of the organization in the eyes of the public is vital for a business to continue. This research strives to continue the discussion regarding how AI is perceived and in what areas the technology would excel. In particular, this study aimed to improve crisis strategies and give organizations options when developing crisis strategies. While AI and communication has a growing amount of research being conducted, this study looked to fill the gap when looking at AI through a crisis communication scope. Businesses need to see what effects a strategy might have on stakeholders before putting a tactic into use during a vulnerable time in the organization.

AI and trust have been heavily studied (Ashoori & Weisz, 2019; Edwards et al., 2014; Jakesch et al., 2019;). In this thesis, the trust of AI was studied in two different areas; the trust of AI as an information source in an organization and the trust of the organization who used the AI Chatbot as a representative. The trust in AI was not significantly different than participant's trust in the human representative. Participants also did not differ in their trust levels of the organization despite which media representative was used. This could be caused by the growing comfortability that people have begun to have in AI as an information source (Ashoori & Weisz, 2019). This is also shown in the fact that the perceived reputation of the organization also did not differ based on which representative was shown.

There was no difference in the perceived credibility of the representative compared to the human representative. Participants were willing to accept that what the representative had to say

about the crisis situation was accurate. Other studies have also found a similar result in credibility when looking at AI (Edwards et al., 2014).

This study also opens the door for many different avenues for further research. Future research could potentially use a variety of crisis types and crisis response strategies in relation to the effectiveness of AI. Research could also investigate other metrics of AI effectiveness in businesses including the sales rates before and after the crisis response or the organizational growth. Finally, future research could also see what characteristics of AI could be manipulated in order to be the most effective within the crisis communication field.

Theoretical Implications

This thesis paired AI with SCCT in an attempt to create a crisis that would invoke a reaction from stakeholders and pair it with a response that could gauge the expected response. The crisis given to participants, human-error product harm, was used as it falls in the preventable cluster and preventable crises create the largest threat to an organization (Coombs, 2007). In response to a preventable crisis, SCCT recommends using a rebuild strategy. The rebuild strategies are best when an organization needs to regain their reputational assets (Coombs, 2007). It could be argued that the strategy accomplished exactly what SCCT lays out. In this study, the reputation of the organization did not differ regardless of who was used to deliver the crisis response. In addition to upholding the reputation of the organization, the SCCT rebuild strategy also maintained the trust in the message source, organization and the credibility despite the unconventional media representative. Historically, SCCT has recommended that someone in a leadership position should deliver the crisis response strategy (Vercic et al., 2018). This leads to the implications that AI should have caused lower levels of trust, credibility, and a decrease in reputation. However, this could be caused by the increasing use of AI in everyday life.

Stakeholders may have started to see AI as an essential part of organizations, thus, putting them into the same level as a human in a leadership role.

SCCT has many moving parts. This study only looked into one preventable crisis situation and one response using a fictitious organization and crisis. As stated above, the theory worked in the way that Coombs anticipated it would. This implies that SCCT is adaptable and can be used in multiple crisis situations and responses even when a secondary component, such as AI, is included. It also gives credit to SCCT that real life organizations may be able to incorporate AI technology into their responses and still be able to navigate the crisis successfully.

Practical Implications

Crisis situations are a very real threat in the everyday life of organization. As organizations prepare for potential crisis, it is important to anticipate what will help navigate a crisis and what might hinder the process. This study has many practical implications. First, this study shows the success of SCCT's rebuild strategies when paired with a crisis that falls into the preventable cluster. A preventable cluster crisis is one of the most damaging crises an organization can have. Being able to successfully navigate through a preventable crisis with little change in the perceived reputation would be considered a large success in any organization, and this study showed that SCCT rebuild strategies has the ability to help with that task.

Another implication of this study is the use of AI within crisis communications. AI has started to be used in many different communication facets including social media (Edwards et al., 2014) online housing rentals (Jakesch et al., 2019) and other large decision-making factors (Ashoori & Weisz, 2019). This thesis paired AI with a crisis response to gauge how stakeholders would react to an AI media representative. Overall, there was little reaction to having an AI

representative in the crisis response. Trust in the organization, organizational reputation and source credibility all had no change regardless of the AI component. These results show that the use of AI can be used in an organizational crisis without hindering the overall success of the organization's crisis response. Using AI instead of a human media representative may allow for organizations to protect their employees if the crisis navigation is not successful, may provide a representative that can respond 24-7 regarding the crisis and potentially allow for communication personnel to continue to monitor the crisis continuously and adjust as needed or even just continue day-to-day business.

Future Directions

SCCT has many moving parts. This thesis only looked at one crisis paired with one response. As a future direction, AI could be paired with each of the response strategies and a crisis that would evoke that response. Not only would this show whether the other response strategies work as they are designed to such as the rebuild strategy did, but also whether AI has an impact on their effectiveness. This direction could also show whether AI is compatible with all crisis types and responses or if stakeholders only respond to the AI representative in some situations.

Another future direction this research could go is including different response delivery methods. In this study, the response was delivered via a press release. In many cases, the response is delivered through many different channels. Some typical crisis responses include press releases, news stories, and a held press conference. A future study could include an AI response given to participants through different channels to see if stakeholders would respond differently to an AI representative if the response was delivered in a different manner.

Finally, this research could also be paired with crisis situations that have already happened. Crisis communication is a field based heavily in case studies. A researcher could take crisis situations that have happened previously and use their response, changing the media representative between the original representative and an AI representative. Using a real crisis situation may be able to invoke a stronger response in stakeholders than a fictitious one. Using a past crisis would also provide a baseline response already from when the crisis happened originally. This response could then be studied against the AI response, and the new human representative response to really see the impacts of the representative. This direction would include the challenge of finding participants who had not been previously exposed to the crisis.

Limitations

As with any research, this study is not without limitations. The first major limitation was the fictitious crisis situation. Using a crisis and an organization that participants had no prior experience with was not able to invoke the same reaction that a real crisis situation would be able to. Because the reaction was likely not as strong as it would be with a real-life crisis, the results are likely tamer than they would be otherwise. Changing the product depicted in the stimuli to one that was more related to the younger participants studied must have also increased the reaction. In addition to using a fictitious organization and crisis, the way the response was delivered could also be a limitation. In this study the crisis response was delivered via a press release with a different media representative presented in the contact information. A press release as a delivery method is already an impersonal method of delivery so using this may have lessened the impact of an AI representative. As a future direction, a different delivery method could be used to see if a more personable delivery method would change the effect of an AI representative.

References

- Ashoori, M., Weisz, J. D. (2019). In ai we trust? Factors that influence trustworthiness of ai-infused decision-making processes. Cornell University.
<https://arxiv.org/pdf/1912.02675.pdf>
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martio, J., Mehta, S., Mojsilovic, A., Nagar, S., Natesan Ramamurthy, K., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney K. R., & Zhang, Y. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. <http://arxiv.org/abs/1810.01943>
- Benoit, W. L. (1995). Accounts, excuses, and apologies: A theory of image restoration strategies. Albany State University. New York Press
- Benoit, W. L. (1997). Image repair discourse and crisis communication. *Public Relations Review*, 23(2), 177-186. [http://dx.doi.org/10.1016/S0363-8111\(97\)90023-0](http://dx.doi.org/10.1016/S0363-8111(97)90023-0)
- Blascovich, J. Social influence within immersive virtual environments. *The social life of avatars: Presence and interaction in shared virtual environments*, 127-145.
- Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, bot or cyborg? *In Proceedings of the 26th Annual Computer Security Applications Conference*, 1, 21-30.
- Claeys A., Cauberghe V., & Vyncke P. (2010). Restoring reputations in times of crisis: An experimental study of the situational crisis communication theory and the moderating effects of locus of control. *Public Relations Review*, 36(3), 256-262.
<https://doi.org/10.1016/j.pubrev.2010.05.004>.
- Coombs, T. W. (1995). Choosing the right words. *Management Communication Quarterly*,

8(4), 447–476. doi:10.1177/0893318995008004003

Coombs, T. W. (2002) Deep and surface threats: Conceptual and practical implications for “crisis” vs. “problem”. *Public Relations Review*, 28(4), 339–345. [https://doi.org/10.1016/S0363-8111\(02\)00167-4](https://doi.org/10.1016/S0363-8111(02)00167-4)

Coombs, T. W. (2004). Impact of past crises on current crisis communication. *Journal of Business Communication*, 41(3), 265-289. <https://doiorg.usd.idm.oclc.org/10.1177/0021943604265607>

Coombs, T. W. (2006). The protective powers of crisis response strategies. *Journal of Promotion Management*, 12(3-4), 241–260. doi:10.1300/j057v12n03_13

Coombs, T. W. (2007) Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corp Reputation Rev*, 10, 163–176. <https://doi-org.usd.idm.oclc.org/10.1057/palgrave.crr.1550049>

Coombs, T. W. (2014). Crisis management and communications. *Institute for Public Relations*. <https://instituteforpr.org/crisis-management-communications/>

Coombs, T. W., & Holladay S. J. (1996). Communication and attributions in a crisis: An experimental study in crisis communication. *Journal of Public Relations Research*, 8(4), 279-295. DOI: 10.1207/s1532754xjpr0804_04

Coombs, T. W., & Holladay, S. (2001). An extended examination of the crisis situations: a fusion of the relational management and symbolic approaches. *Journal of Public Relations Research*, 13(4), 321-340. DOI 10.1207/S1532754XJPRR1304_03

Coombs, T. W., & Holladay, S. J. (2002). Helping Crisis Managers Protect Reputational Assets. *Management Communication Quarterly*, 16(2), 165–186. doi:10.1177/089331802237233

- Coombs, T. W., & Holladay, S. J. (2004). Reasoned action in crisis communication: An attribution theory-based approach to crisis management. In D. P. Millar & R. L. Heath (Eds.), *Responding to crisis: A rhetorical approach to crisis communication* (pp. 95-115). Mahwah, NJ: Lawrence Erlbaum.
- Dean, D. H. (2004). Consumer reaction to negative publicity. *Journal of Business Communication, 41*(2), 192–211.
<https://doi.org.usd.idm.oclc.org/10.1177/0021943603261748>
- Dhankar, M., & Walia, N. (2020). An introduction to artificial intelligence. *Proceedings of 'Emerging Trends in Big Data IoT and Cyber Security*. 105-108.
<http://msi-ggsip.org/wp-content/uploads/conference2020.pdf#page=118>
- Duffy, P. J., Shiflett, S., & Downey, R. G. (1977). Locus of control: Dimensionality and predictability using Likert scales. *Journal of Applied Psychology, 62*(2), 214–219.
<https://doi-org.usd.idm.oclc.org/10.1037/0021-9010.62.2.214>
- Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior, 33*, 372–376. doi:10.1016/j.chb.2013.08.013
- Fan, X., McNeese, M., & Yen, J. (2010). NDM-Based Cognitive Agents for Supporting Decision-Making Teams. *Human-Computer Interaction, 25*(3), 195–234.
doi:10.1080/07370020903586720
- Feurer, M., Klein, A., Eggenberger, K., Springenberg, J. T., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. *Advances in Neural Information Processing Systems. 29*62–2970.

- Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly*, 77(3), 515–540. doi:10.1177/107769900007700304
- Fombrun, C. J., Gardberg, N. A. & Sever, J. M. (2000). The Reputation Quotient: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management*, 7(4), 241–255. doi:10.1057/bm.2000.10
- Fox, J., Ahn, S., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N., (2014). Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Human-Computer Interaction*, 30(5), 401–432. doi:10.1080/07370024.2014.921494
- Griffin, M., Babin, B., & Attaway, J. (1991). An empirical-investigation of the impact of negative public publicity on consumer attitudes and intentions. *Advances in Consumer Research*, 18, 334-341.
- <http://web.a.ebscohost.com.usd.idm.oclc.org/ehost/detail/detail?vid=0&sid=52c0f791-85fd-492c-ad4f-7fdfeba6eb5f%40sdc-v-sessmgr03&bdata=JnNpdGU9ZWwhvc3QtbGl2ZSZzY29wZT1zaXRl#AN=6522208&db=keh>
- Gunkel, D. J. (2017). Communication and artificial intelligence: New opportunities and challenges for communication research. *Galáxia*, 34.
- http://www.scielo.br/scielo.php?script=sci_arttext&pid=S198225532017000100005&lng=pt&tlng=pt
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. ArXivpreprint arXiv:1812.04608.
- Hon, L. C., & Grunig, J. E. (1999) Guidelines for measuring relationships in public relations.

Institute for Public Relations. https://www.instituteforpr.org/wp-content/uploads/Guidelines_Measuring_Relationships.pdf

- Jakesch, M., French, M., Ma, X., Hancock, J. T., & Naaman, M. (2019). AI-Mediated Communication. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. doi:10.1145/3290605.3300469
- Kim, T., & Song, H. (2020). How should intelligent agents apologize to restore trust?: The interaction effect between anthropomorphism and apology attribution on trust repair. *PsyArXiv*. <https://doi.org/10.31234/osf.io/234f5>.
- Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, 68, 57–68. doi:10.1080/03637750903074685
- Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. *Australasian conference on information systems*, 53., 6–8. <http://citeseerx.ist.psu.edu/viewdoc/summary>
doi=10.1.1.93.3874
- McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs*, 41(3), 261-266. <https://doi.org/10.1080/03637757409375845>
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, 66(1), 90-103. <https://doi.org/10.1080/03637759909376464>
- McDowell, Z. J., & Gunkel, D. J. (2016). Introduction to machine communication. *Communication +1*, 5(1). <https://scholarworks.umass.edu/cpo/vol5/iss1/1>
- Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in

- the contemporary media environment. *Annals of the International Communication Association*, 27(1), 293–335. doi:10.1080/23808985.2003.1167902
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.
<https://doi-org.usd.idm.oclc.org/10.1016/j.artint.2018.07.007>
- Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012) Tweeting is believing? Understanding microblog credibility perceptions. *In Proceedings of the ACM 201 conference on Computer Supported Cooperative, Work (CSCW2012)*, 441-450.
- Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, 432-440.
<https://doi.org/10.1016/j.chb.2017.02.067>.
- Murawski, J. (2019). Mortgage providers look to AI to process home loans faster. *Wall Street Journal*. <https://www.wsj.com/articles/mortgage-providers-look-to-ai-to-process-home-loans-faster-11552899212>
- Nah, S., McNealy, J., Hyun Kim, J. & Joo, J. (2020) Communicating artificial intelligence (AI): Theory, research, and practice. *Communication Studies*, 71(3), 369-372. DOI: 10.1080/10510974.2020.1788909
- Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2008). The mere belief in social interaction improves learning. *Proceedings of the ICSL 2008 International Conference on International Conference for the Learning Sciences*. Mahwah, NJ: Erlbaum.
- Pauly, J., & Hutchison, L. (2005). Moral fables of public relations practice: The Tylenol and Exxon Valdez cases. *Journal of Mass Media Ethics*, 20 (4), 231–249.
https://doi.org/10.1207/s15327728jmme2004_2

- Rudin, C. (2018). Please stop explaining black box models for high stakes decisions. ArXiv preprint arXiv:1811.10154
- Sangwon Lee, Seungahn Nah, Deborah S. Chung & Junghwan Kim (2020) Predicting AI news credibility: Communicative or social capital or both?, *Communication Studies*, 71(3), 428-447. DOI: 10.1080/10510974.2020.1779769
- Seeger, W. M., Sellnow, L. T., & Ulmer, R. R. (2003). *Communication and organizational crisis*. Westport, CT: Greenwood Publishing Group.
- Sellnow, L. T., & Seeger W. M. (2013). *Theorizing crisis communication*. Wiley-Blackwell.
- Sisco, H. F. (2012). Nonprofit in crisis: An examination of the applicability of situational crisis communication theory. *Journal of Public Relations Research*, 24(1), 1–17. doi:10.1080/1062726x.2011.5822
- Song, H., McComas, K. A., & Schuler, K. L. (2018). Source effects on psychological reactance to regulatory policies: The role of trust and similarity. *Science Communication*, 107554701879129. doi:10.1177/1075547018791293
- Spitzberg, B. H. (2006). Preliminary development of a model and measure of computer-mediated communication (CMC) competence. *Journal of Computer-Mediated Communication*, 11(2), 629-666. <https://doi.org/10.1111/j.1083-6101.2006.00030.x>
- Verčič, A. T., Verčič, D., & Coombs, T. W. (2018). Convergence of crisis response strategy and source credibility: Who can you trust? *Journal of Contingencies and Crisis Management*. doi:10.1111/1468-5973.12229
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Department of Psychology, University of California*, 92(4), 548-573. 10.1037/0033-295X.92.4.548
- Westerman, D. W., Spence, P.R., Van Der Heide, B. (2012). A social network as information:

The effect of system generated reports of connectedness on credibility and health care information on Twitter. *Computers in Human Behavior*, 28(1), 199-

206. <https://doi.org/10.1016/j.chb.2011.09.001>

Zhao, D., & Rosson, M.B. (2009). How and why people Twitter: The role that microblogging plays in informal communication at work. *In Proceedings of the SIGCHI International Conference on Supporting Group, Work (ACM2009)*, 2009. pp. 243– 252.

Appendix A

Message Source Trust t-Test

Table A1

Message Source Trust t-Test

	<i>AI</i>	<i>Hum</i>
Mean	3.22310606	3.4829932
Variance	0.70126074	0.65778339
Observations	88	98
Hypothesized	0	
df	180	
t Stat	-2.1449034	
P(T<=t) one-t	0.01665061	
t Critical one-	1.65336301	
P(T<=t) two-t	0.03330123	
t Critical two-	1.97323082	

Appendix B

Organizational Trust t-Test

Table B1

Organization Trust t-Test

	<i>AI</i>	<i>Hum</i>
Mean	3.46022727	3.39030612
Variance	1.89064133	0.58384836
Observations	88	98
Hypothesized	0	
df	133	
t Stat	0.42208423	
P(T<=t) one-t	0.33682252	
t Critical one-	1.65639124	
P(T<=t) two-t	0.67364504	
t Critical two-	1.97796126	

Appendix C

Organizational Reputation t-Test

Table C1

Organizational Reputation t-Test

	<i>AI</i>	<i>Hum</i>
Mean	3.24090909	3.24081633
Variance	0.29807732	0.21604881
Observations	88	98
Hypothesized	0	
df	172	
t Stat	0.00124052	
P(T<=t) one-t	0.49950582	
t Critical one-	1.65376095	
P(T<=t) two-t	0.99901164	
t Critical two-	1.97385217	

Appendix D

Credibility t-Test

Table D1

Credibility t-Test

	<i>AI</i>	<i>Hum</i>
Mean	2.94090909	2.91836735
Variance	0.50681296	0.44357669
Observations	88	98
Hypothesized	0	
df	179	
t Stat	0.22226657	
P(T<=t) one-t	0.41217988	
t Critical one-	1.6534108	
P(T<=t) two-t	0.82435977	
t Critical two-	1.97330543	

Appendix E

Stimuli

Jacobson Corporation has been the leading vitamin company on the east coast for the last 50 years. Recently, the Jacobson Co. has taken a hit in sales due to a string of employee misconduct complaints that have been made public and have resulted in the firing of two individuals.

On July 10, 2020, the company sent out its regular distribution of gummy multivitamins to local drug stores. One week after this specific shipment hit the shelves, consumers began to report feeling ill after ingesting the multivitamin. Jacobson Co. heard the complaints and immediately pulled the vitamins off the shelves and launched an investigation. This investigation revealed that the vitamins within the shipment had been tampered with by an employee of the company.

As a result of the tampering, 50 consumers were hospitalized from ingesting the multivitamin. The Jacobson Corporation's public reputation and sales took a major hit because of this crisis. As a result, the corporation released the following crisis response delivered via a press release created by Jacobson Co.'s media representative, Pat Kelly/ChatBot AI Agent.

Appendix F

Human Response

Contact:

Pat Kelly

Media Representative for Jacobson Co.

mediarelations@jacobson.org

PRESS RELEASE

FOR IMMEDIATE RELEASE

Jacobson Corporation Concludes Investigation into Multivitamin Tampering

Following the investigation of the 7-10-20 multivitamin shipments, our team has found that a Jacobson employee had tampered with products to intentionally cause harm to the public and to the corporation.

Jacobson Corporation understands the negative reflection this one employee has cast onto the corporation as a whole. As a result, the company is prepared to take full responsibility for the product tampering that has jeopardized many lives.

The corporation has let go of the employee in question and has committed to reviewing all protocols regarding product quality control to ensure this situation will not happen again. Jacobson Corporation would like to extend a sincere apology to all the families affected by this crisis.

Any further questions or concerns can be directed to mediarelations@jacobson.org

###

Appendix G

AI Response

Contact:

ChatBot Artificial Intelligence (AI) Agent

Media Representative for Jacobson Co.

mediarelations@jacobson.org

PRESS RELEASE

FOR IMMEDIATE RELEASE

Jacobson Corporation Concludes Investigation into Multivitamin Tampering

Following the investigation of the 7-10-20 multivitamin shipments, our team has found that a Jacobson employee had tampered with products to intentionally cause harm to the public and to the corporation.

Jacobson Corporation understands the negative reflection this one employee has cast onto the corporation as a whole. As a result, the company is prepared to take full responsibility for the product tampering that has jeopardized many lives.

The corporation has let go of the employee in question and has committed to reviewing all protocols regarding product quality control to ensure this situation will not happen again. Jacobson Corporation would like to extend a sincere apology to all the families affected by this crisis.

Any further questions or concerns can be directed to mediarelations@jacobson.org

###

Appendix H

Survey Measures

Message Source Trust (Song et al, 2018)

On a scale of 1 to 5, 1 being “strongly disagree”, 3 being “neither disagree nor agree” and 5 being “strongly agree”, rate your agreement with the following statements.

I trust the media representative to...

Provide the best available information on the crisis

Provide enough information to allow me to decide my stance on the crisis

Provide truthful information about human safety regarding the crisis

Provide timely information regarding the crisis

I feel that the media representative...

Shares similar values as me

Shares similar opinions as me

Organization Trust (Hon & Grunig, 1999)

On a scale of 1 to 5, 1 being “strongly disagree”, 3 being “neither disagree nor agree” and 5 being “strongly agree”, rate your agreement with the following statements.

Jacobson Corporation...

Can be relied on to keep its promises

Treats people like me fairly

Has the ability to accomplish what it says it will do

I feel confident about Jacobson Corporation's skills to overcome this crisis

Organizational Reputation (Coombs & Holladay, 2002)

On a scale of 1 to 5, 1 being "strongly disagree", 3 being "neither disagree nor agree" and 5 being "strongly agree", rate your agreement with the following statements.

Jacobson Corporation is concerned with the well-being of its publics

Jacobson Corporation is basically dishonest

I do not trust Jacobson Corporation to tell the truth about the incident

Under most circumstances I would be likely to believe what Jacobson Corporation says

Jacobson Corporation is not concerned with the well-being of its publics

Source Credibility (Metzger et al, 2003)

On a scale of 1 to 5, 1 being "none at all", 3 being "a moderate amount" and 5 being "a great deal", rate your agreement with the following statements.

To what degree do you find the media representative to be...

Believable

Accurate

Trustworthy

Biased

Thorough in the information provided