

Course: SKOM12
Term: Spring 2026
Supervisor: Howard Nothhaft
Examiner: Mats Heide

Navigating the Grey Zone: Estonian Communication Practitioners' Perspectives on Ethical AI Use and Responsibility in Crisis Communication

KARMEN KIKAS

Lund University
Department of Strategic Communication
Master's Thesis



Abstract

Navigating the Grey Zone: Estonian Communication Practitioners' Perspectives on Ethical AI Use and Responsibility in Crisis Communication

Artificial intelligence (AI) is increasingly employed in crisis communication to assist with tasks like monitoring, analysis, and message creation. However, its expanding use raises ethical and responsibility concerns, especially when AI-supported communication results in problematic outcomes. This thesis aims to provide new knowledge and enhance the understanding of how Estonian communication professionals perceive the ethical use of AI and responsibility in crisis communication. Using semi-structured, vignette-based interviews with 16 practitioners from 13 organisations and drawing on moral responsibility and machine ethics, the study explores how practitioners define AI's role, where they set boundaries between human and AI tasks, and how they consider responsibility when AI-related errors occur. The analysis indicates that practitioners view AI as a support tool but recognise that it can subtly influence their interpretation of crises. Factors such as crisis severity, the ability to verify outputs, and data sensitivity primarily guide the division of tasks between humans and AI. Regarding moral responsibility, practitioners do not see it as a single issue: crisis conditions can diminish judgements of individual competence, organisational hierarchy can influence who is held answerable for explaining failures regardless of who used the AI, and accountability for outcomes remains unclear. The thesis highlights that practitioners using AI operate in a highly uncertain environment, where few widely agreed-upon norms or established practices guide their actions. It therefore underscores the need for organisational AI guidelines and structured evaluation procedures to better support practitioners, who currently face AI-related challenges largely on their own.

Keywords: crisis communication, artificial intelligence, AI, machine ethics, moral responsibility, attributability, answerability, accountability, responsibility gap

Word count: 19,986

Acknowledgements

I sincerely thank my supervisor, Howard Nothhaft, for his guidance and support during the writing process. His expert advice and insightful feedback greatly contributed to the development of this thesis. I am also very grateful to the 16 Estonian communication professionals who took part in this thesis and generously shared their valuable insights.

Table of Contents

1. Introduction	1
1.1. Aim and Research Questions	3
1.2. Delimitations	4
2. Literature Review	6
2.1. Defining AI	6
2.1.1. Types of AI	7
2.2. AI in Crisis Communication	8
2.2.1. Defining Crisis Communication	9
2.2.2. Use of AI Tools in Crisis Communication	10
2.2.3. Ethical Considerations Regarding AI in Crisis Communication	12
2.2.4. Human Versus AI-led Communication Tasks	13
2.3. Responsibility and AI in Crisis Communication	14
2.4. Synthesis	17
3. Theory	18
3.1. Moral Responsibility and its Three “Faces”	18
3.2. Machine Ethics and Responsibility	21
3.2.1. AI and Responsibility Gap	23
4. Methodology	25
4.1. Research Paradigm	25
4.2. Research Design	26
4.3. Collection of Empirical Material and Interview Proceedings	27
4.3.1. Selection of Participants	28
4.4. Analysis of Empirical Material	30
4.5. Ethical Considerations and Reflection	31
5. Analysis	33
5.1. Use of AI in Crisis Communication	33
5.1.1. The Routine Work with AI: Efficiency Versus Loss of Skills	33
5.1.2. AI as Sensemaking Support, but Not a Sensemaker	35

5.1.3. The “Black Box” Problem.....	37
5.1.4. Governing AI Through (Absent) Guidelines.....	38
5.2. Task Division Between Human and AI	40
5.2.1. Crisis Severity as Decision Threshold.....	40
5.2.2. Verification Ability	42
5.2.3. Data Sensitivity	43
5.3. Understanding Responsibility with AI-related Errors	45
5.3.1. Attributability: Crisis Softens the Judgement of Actions.....	45
5.3.2. Answerability: Managers in Charge.....	47
5.3.3. Accountability: The “Non-Negotiable” Human Oversight.....	48
6. Concluding Discussion	51
6.1. Contributions to Research and Practice	55
6.2. Recommendations for Future Research.....	56
References	58
Appendix 1. Interview Guide	69
Appendix 2. Consent Form.....	73

1. Introduction

On February 13, 2023, a tragic mass shooting at Michigan State University in East Lansing, Michigan, resulted in three students losing their lives and five others being seriously injured. In the days that followed, Vanderbilt University sent a message to its students expressing sympathy for the tragedy. When it was later revealed that this comforting email was generated using OpenAI's ChatGPT, it sparked widespread criticism and a debate over the ethics of using AI-generated communication during times of collective grief (Levine, 2023; Piller, 2025, p. 43). Consequently, this incident underscores a major challenge for crisis communication professionals: when utilising AI to craft quick and emotionally sensitive messages, how should they determine what is appropriate, and who bears responsibility if the communication causes controversy or worsens the situation?

Regarding ethical crisis communication, the problem in this school shooting example is not simply that AI was used, but how the issues of appropriateness and responsibility towards affected audiences were handled. Ethical crisis communication involves organisations prioritising messages that show care, empathy, and sensitivity toward those affected by the crisis, rather than emphasising efficiency or protecting organisational reputation (Falkheimer, 2021, p. 6). Seen through Coombs' (1995, p. 455) concept of a *faux pas*¹, the criticism of the AI-generated message reflects how a communication initially considered appropriate by the organisation was reinterpreted as inappropriate by external audiences. The audience felt the message lacked empathy during a moment of shared grief, raising concerns about professionalism and ethics.

AI tools have become integrated into communication practices and are now widely used, including in crisis communication (Panda et al., 2019, p. 204). The primary benefit of using AI in crisis communication is its independence from human limitations: it does not sleep or eat and is unaffected by routine bodily needs that

¹ A *faux pas* refers to a situation in which an organisation's unintentional and well-intended action is redefined by external audiences as inappropriate, creating ambiguity and transforming the action into a crisis (Coombs, 1995, p. 455).

typically interrupt or delay work (Nies & Zhao, 2025, p. 121). This allows AI to operate continuously, supporting rapid monitoring and detection of emerging issues across multiple sources even outside normal working hours (Nies & Zhao, 2025, p. 121). However, as with many things, the good rarely exists without the bad. Volk and Buhmann (2023, pp. 128–129) highlight that while AI can support strategic communication in various ways, it also creates new problems and dilemmas, such as producing misleading results and making it difficult to understand or explain how decisions are made. Furthermore, its use raises concerns about transparency and responsibility, as practitioners must evaluate its reliability and justify decisions made with its support (Volk & Buhmann, 2023, p. 128). This is especially relevant in crisis communication, a key part of strategic communication, where organisations need guidance on how to respond publicly under pressure and uncertainty (Coombs & Holladay, 2022, p. 259). Strategic communication, in this sense, is the purposeful use of communication to achieve organisational goals (Hallahan et al., 2007, p. 3). It extends beyond drafting messages: professionals serve as strategic advisors who analyse situations, predict stakeholder reactions, and guide decisions in high-stakes, unpredictable scenarios (Falkheimer et al., 2016, p. 150; Zerfass et al., 2018, p. 495).

Introducing AI tools into these tasks begins to shape the communication process by affecting which information is prioritised and how messages are created (Gunkel, 2012, p. 23). AI no longer acts just as a technical channel but becomes an embedded part of strategic communication (Gunkel, 2012, p. 23). As AI becomes more integrated into everyday communication work, the field faces uncertainty because routine use outpaces the development of norms and institutional frameworks that should guide the implementation of AI (Shad, 2025, pp. 891–892). This change raises ethical questions about its use, since professionals remain responsible for the advice they give and the messages they issue on behalf of their organisation (Steyn, 2003, p. 178). When AI-supported communication results in misinformation or harm during crises, responsibility cannot be assigned solely to the technology but must be considered alongside professional judgement and organisational decisions (Schoenherr & Thomson, 2024, p. 66). Therefore, understanding how practitioners reason about the ethical use of AI and responsibility is crucial for strategic crisis communication, as it highlights how they justify decisions, handle ethical issues, and navigate responsibility in high-stakes situations.

Estonia is a particularly relevant context for exploring these questions, given its high level of digitalisation and early adoption of advanced technologies across sectors (Trapp, 2025). The country has a nationally led AI initiative focused on increasing the value of AI across national sectors, and the use of AI to enhance work efficiency is strongly encouraged (Government Office of Estonia, n.d.). Thus, by focusing on Estonian communication professionals with significant practical experience using AI tools, this study sheds light on how they manage responsibility and ethical challenges in crisis communication. Instead of labelling AI as good or bad, this thesis examines how practitioners themselves evaluate its appropriateness, handle problematic results, and assign responsibility.

1.1. Aim and Research Questions

The aim of this thesis is to provide new knowledge and enhance the understanding of how communication professionals in Estonia understand and negotiate questions of ethical AI use and responsibility when using AI tools in crisis communication. Specifically, the study explores how practitioners see where AI adds value or limits the work, how they distinguish between human and AI responsibilities, and how they conceptualise responsibility when AI-supported communication contributes to problematic or unintended outcomes in high-paced situations. The empirical focus is on practitioners who already use AI tools in their everyday work, ensuring that participants' perspectives are grounded in practical experience rather than speculation. Grounded in this aim, the following research questions are formulated:

RQ1: How do Estonian communication professionals describe the use of AI in crisis communication?

RQ2: How do Estonian communication professionals decide which crisis communication tasks are appropriate to assign to AI and which must remain human-led?

RQ3: When AI-related errors occur in crisis communication, how do Estonian communication professionals determine who is responsible?

The research questions are designed to examine the topic across different stages of professional practice. The first question explores how AI is integrated into everyday crisis communication work, establishing the practical context in which decisions

about ethics and responsibility are made. The second question moves beyond general descriptions of AI use and focuses on how practitioners draw boundaries between AI-supported and human-led tasks. By examining which crisis communication tasks professionals trust AI to handle and which they believe require human judgement, this question shows how practitioners assess AI's capabilities and limitations in high-stakes situations. The third question then explicitly addresses responsibility by investigating how communication professionals interpret and assign moral responsibility when AI-related errors occur in crisis communication.

Through these three research questions, the thesis contributes to the strategic communication field by providing new knowledge about how communication practitioners negotiate the ethical dilemmas and responsibility ambiguity when using AI. As mentioned, AI has created considerable uncertainty across sectors (Shad, 2025, pp. 891–892), and strategic communication, too, is entering a phase in which its professional boundaries are being renegotiated (Bowen & Erzikova, 2025, p. 3). Crises intensify the core demands of strategic communication, such as shaping stakeholder understanding and behaviour, maintaining legitimacy and trust, protecting reputation, and ensuring efficiency (Falkheimer & Heide, 2022, p. 8). Since AI will increasingly become part of this work, its role, as well as its ethical use, will become more pressing issues for strategic communication than ever before.

1.2. Delimitations

The thesis has several delimitations. Firstly, it should be mentioned that it focuses only on communication practitioners who incorporate AI into their daily tasks and have experience in crisis communication. Other relevant parties, such as top management of organisations, AI developers, legal experts or affected stakeholders, are excluded because the focus is on understanding how ethical AI use and responsibility are perceived within *communication* practices. Additionally, the study uses a fictional crisis vignette rather than a real case to prompt practitioners' reflections in interviews. While the vignette helps to protect confidentiality, it captures the practitioners' expected reasoning rather than their responses under actual crisis conditions. In real situations, consequences and emotions might influence judgement differently. Lastly, it should be considered that AI is a rapidly evolving field. The thesis and its empirical material reflect a specific period of time (February-May 2026) in AI's integration into

crisis communication. Future advances in tools and changes in AI governance can influence both AI applications and perceptions of responsibility. As a result, some conditions and arguments discussed in the thesis are likely to evolve, and new ethical and practical issues will arise. The analysis should be read with this temporal context in mind.

2. Literature Review

This chapter presents previous research on AI in strategic and crisis communication, with attention to its ethical boundaries and implications for responsibility. The literature review begins by examining key definitions of AI to clarify how it is understood in this thesis. Establishing a clear understanding of AI is necessary, as it is a multidimensional concept defined differently across disciplines. The chapter then reviews current research on the use of AI in crisis communication practice. Finally, it discusses the issues around responsibility and ethics to give an overview of the risks and wider impacts of using AI in crises.

2.1. Defining AI

The definition of AI has been the subject of ongoing scholarly debate, as there is no widely unified and agreed-upon definition (Ertel, 2017, p. 1; Kaplan, 2021, p. 22; Wang, 2019, p. 3). AI is not a recent phenomenon, as experiments involving human-like machines already took place in the early 20th century (Ertel, 2017, p. 6). The roots of AI trace back to 1940s science fiction, where the “Three Laws of Robotics” were introduced, shaping early ideas about intelligent machines (Haenlein & Kaplan, 2019, p. 6). Simultaneously, Alan Turing laid the theoretical groundwork for machine intelligence, including the Turing Test (Haenlein & Kaplan, 2019, pp. 6–7; Muggleton, 2014, p. 4). The field was formally established in the 1950s at the Dartmouth Summer Research Project on Artificial Intelligence (Haenlein & Kaplan, 2019, p. 7).

The definitions of AI differ depending on which aspects of intelligence are highlighted. Russell and Norvig (2023, p. 20) mention that AI can be described as algorithms, human-like reasoning, specific skills, or systems that operate within their environment. All these perspectives suggest that AI involves systems that interpret their surroundings and act autonomously to achieve goals (Russell & Norvig, 2023, p. 20). Franklin (2014, p. 15) identifies two main categories of AI: an engineering approach focused on creating systems that meet human needs, and a scientific approach aimed at understanding human intelligence through computational models.

Building upon this, Kaplan (2021, p. 23) classifies AI into analytical, human-inspired, and humanised systems, depending on how closely they mimic cognitive, emotional, and social intelligence.

Nevertheless, McCarthy and Hayes (1981, p. 431) stress that for a computer program to act *intelligently*, it requires a generalised representation of the world that enables it to make sense of its inputs. They therefore reject a purely behaviourist view of AI and instead define it in terms of internal world models (McCarthy & Hayes, 1981, p. 432). An intelligent system, on their account, can answer questions, predict outcomes, acquire new information when needed, and act in accordance with its goals (McCarthy & Hayes, 1981, p. 432). Ertel (2017, p. 2) likewise rejects purely behaviour-based definitions of AI, since they also count simple systems as intelligent, and instead follows Rich (1985, p. 117) in defining AI as the field concerned with making computers perform tasks at which humans are currently better (e.g., problem solving, language understanding). This means that AI aims to solve real-world problems, many of which involve people and culture (Rich, 1985, p. 117).

In this thesis, AI is understood as systems that perform human-like tasks by using internal models to interpret information, reason, and act purposefully (Ertel, 2017, p. 2; Franklin, 2014, p. 15). This definition is relevant to strategic crisis communication because AI can be understood as more than a technical system. It functions as a communicative agent that shapes interactions and influences communication, rather than simply carrying out tasks (Guzman & Lewis, 2020, pp. 71–72).

2.1.1. Types of AI

Before exploring AI's role in crisis communication, it is essential to understand the types of AI currently used in communication tasks. Sarker (2022, p. 3) argues that AI encompasses a broad range of systems designed to handle tasks traditionally performed by humans, such as learning or reasoning. Based on this, AI can be divided into five main categories: analytical, functional, interactive, textual, and visual AI (Sarker, 2022, p. 3). Analytical AI is used for data analysis to inform decision-making; functional AI performs actions based on data; interactive AI enables automated user communication; textual AI processes and generates human language; visual AI interprets images and videos to extract meaning (Sarker, 2022, pp. 3–4).

Building on this functional classification, Rashid and Kausik (2024, p. 5) further differentiate AI systems by capability, distinguishing three levels: narrow AI, general (or strong) AI, and super AI. Narrow AI is designed to perform specific, well-defined tasks within limited domains, and it lacks general intelligence beyond its programmed functions (Damar et al., 2024, p. 87; Rashid & Kausik, 2024, p. 6). All existing AI systems, including those used in communication, fall into this category, as they are designed for specialised tasks and are considered weak AI (Bory et al., 2024, p. 2107). Despite this label, they can still significantly influence everyday practices (Bory et al., 2024, p. 2109). In contrast, general or strong AI refers to a theoretical form of AI capable of performing any intellectual task a human can, including the ability to transfer knowledge and skills across contexts (Goertzel, 2014, p. 3). Super AI represents an even more advanced, human-like intelligence that would surpass human cognitive abilities in all domains, including reasoning, creativity, and decision-making (Rashid & Kausik, 2024, p. 7). Nevertheless, both general and super AI remain hypothetical, as such highly intelligent systems do not yet exist and largely belong to the realm of scientific speculation (Bory et al., 2024, p. 2107).

2.2. AI in Crisis Communication

Research indicates that AI is already embedded in many daily communication activities within strategic communication (Buhmann et al., 2026, p. 3; Gregory & Smith, 2025, p. 122; Nies & Zhao, 2025, p. 120). The European Communication Monitor study by Zerfass et al. (2024, p. 26) shows AI is most prevalent in monitoring, analysis, content adaptation, and inspiration gathering. Other tasks, such as stakeholder communication or managing internal assets like contacts, show lower adoption levels (Zerfass et al., 2024, p. 26). Similarly, Gregory and Smith's (2025, p. 122) AI-task map reveals that tools are mainly used for content-related activities, including content creation, optimisation, social media management, and cross-platform monitoring. AI also significantly contributes to data analysis, social listening, and stakeholder sentiment tracking (Gregory & Smith, 2025, p. 122). Overall, AI supports key communicative functions like information monitoring, processing, and analysis (Buhmann et al., 2026, p. 3). These functions are especially vital during periods of high uncertainty and limited information, such as in crisis communication (Sellnow & Seeger, 2013, p. 7).

2.2.1. Defining Crisis Communication

A crisis is commonly understood as an unexpected event that disrupts an organisation's operations and endangers its reputation, financial health, and relationships with stakeholders (Coombs, 2007, p. 164). A key feature of a crisis is the need for immediate response to limit and control potential damage (Sellnow & Seeger, 2013, p. 7). Coombs (2006, p. 243) notes that the severity of a crisis depends on the level of financial, physical, environmental, or emotional damage it causes. Since crises are unpredictable and demand quick action, effective communication is vital for guiding responses and reducing uncertainty, as stakeholders require clear, accurate information about what occurred, what is known, and what measures are being taken (Coombs, 2007, p. 165). When an organisation's actions are perceived as inappropriate, crisis communication shifts from mere explanation to restoring credibility and safeguarding its public image (Hearit, 2006, p. 12).

However, crisis communication should not be understood as a single, largely informational task. Sturges (1994, p. 299) cautions against such a static view and, instead, conceptualises crises as processes that unfold over time. Drawing on the idea of a crisis life cycle, he argues that crises develop through stages, beginning with early warning signs ("crisis build up"), when the first hints of a crisis appear (Sturges, 1994, p. 299). The second stage is "crisis breakout", during which an exceptional event triggers a crisis (Sturges, 1994, p. 300). The third phase, "abatement", is the period after the acute crisis when its effects continue to unfold over time, requiring organisations to manage ongoing consequences and follow-up actions until the crisis is resolved (Sturges, 1994, p. 300). Finally, "termination" marks the point at which the crisis is considered fully resolved and no longer poses a concern for the organisation or its stakeholders (Sturges, 1994, p. 300).

Frandsen and Johansen (2011, p. 355) divide crisis stages more broadly: 1) the pre-crisis phase, which focuses on prevention and preparation for potential risks, 2) the crisis phase, which centres on managing the situation and making sense of unfolding events, and 3) the post-crisis phase, which supports learning and the implementation of necessary changes. As a crisis moves through these phases, communication tasks such as information gathering, message dissemination, and stakeholder engagement also change as organisations face different pressures and stakeholder expectations at each stage (Sturges, 1994, p. 300). Thus, crisis communication can be seen as a sensemaking process in which organisational actors interpret events and

negotiate shared understandings (Zhao et al., 2017, pp. 367–368). As Weick (1988, p. 305) implies, sensemaking becomes particularly important in crises because unexpected events disrupt existing interpretations and compel actors to understand what is happening as the situation unfolds. Crises impose severe demands on sensemaking because people must interpret events and act under conditions of uncertainty and time pressure (Weick, 1988, p. 305). In this light, AI tools are relevant as they can help professionals organise information, explore interpretations, and thus better understand situations before deciding how to respond (Cheng et al., 2024, p. 176).

The phases model provides a helpful framework for understanding how communication tasks develop over time, but it has been criticised for depicting crises as too linear and distinctly separated (Frandsen & Johansen, 2011, p. 354). Jaques (2007, p. 150) argues that this approach may overlook overlaps between phases, ongoing uncertainty, and the reality of organisations facing multiple crises simultaneously. Instead, crises should be viewed as an integrated, non-linear process in which preparation, prevention, response, and post-crisis learning continuously overlap and influence one another (Jaques, 2007, p. 150). Still, Frandsen and Johansen (2011, p. 354) maintain that the staged model remains a valuable tool for organising and analysing crisis communication, without suggesting crises occur in strictly defined stages. This literature review also examines AI in crisis communication through the pre-crisis, crisis, and post-crisis phases, as this framework helps organise existing research while recognising these phases often overlap in practice (Cheng et al., 2024, p. 176; Frandsen & Johansen, 2011, p. 354; Jaques, 2007, p. 150). Specific examples of AI use in different crisis phases are discussed in the next subchapter.

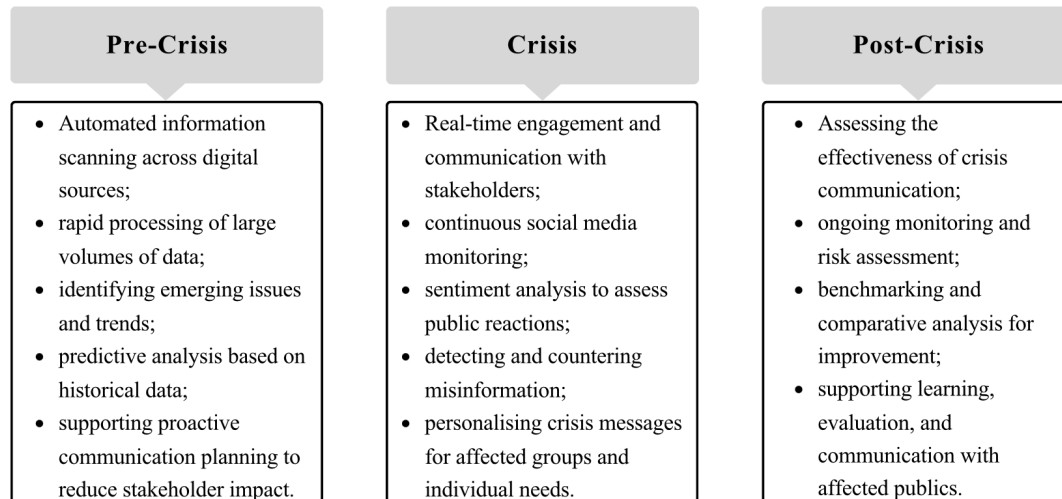
2.2.2. Use of AI Tools in Crisis Communication

Crises are fast-paced and time-pressured, requiring organisations to process large amounts of information and communicate quickly with diverse stakeholders (Cheng et al., 2024, p. 173). In response, practitioners increasingly use AI tools to support information processing, message sharing, and interaction with affected stakeholders (Cheng et al., 2024, p. 177). From a communication perspective, Cheng et al. (2024, p. 176) see AI as a supportive tool that shapes how organisations collect, interpret, and disseminate information across different stages of a crisis, as shown in Figure 1,

which corresponds to the pre-crisis, crisis, and post-crisis phases (Frandsen & Johansen, 2011, p. 355).

Figure 1

Functions of AI in Different Crisis Phases



Note. Functions of AI in different crisis phases. Adapted from “Crisis Communication in the Age of AI: Navigating Opportunities, Challenges, and Future Horizons”, by Y. Cheng, J. Lee, and J. Qiao, in W. T. Coombs (Ed.), *Media and Crisis Communication* (1st ed., p. 176), 2024, Routledge (<https://doi.org/10.4324/9781032619187>). Copyright 2024 by Routledge.

In crisis communication, AI is employed to predict risks, facilitate real-time messaging, and assess responses afterwards (Cheng et al., 2024, p. 176). It also helps reduce repetitive tasks, allowing human resources to focus on more complex duties (Cheng et al., 2024, p. 177). Buhmann et al. (2026, p. 3) highlight AI’s role in media and sentiment analysis, noting its capacity to process large volumes of data to better detect tone and emotions (Nies & Zhao, 2025, p. 122). Hossain et al. (2025, pp. 155, 157) describe AI as a decision-support tool that processes real-time data, enhances assessments and response planning, and helps with information sharing during crises. Hancock et al. (2020, p. 91) also demonstrate that AI can support communication by correcting spellings, suggesting replies, generating full messages, and enabling quick interactions via text, audio, and video.

2.2.3. *Ethical Considerations Regarding AI in Crisis Communication*

AI can enhance the speed and efficiency of crisis communication, but it also raises ethical and moral questions about responsibility and trust. Panda et al. (2019, p. 204) highlight a distinct professional boundary: practitioners appreciate AI for issue detection, managing large data sets, and post-crisis evaluations, yet they firmly oppose using it to handle a crisis independently. This resistance stems from concerns about client acceptance and the necessity of human judgement in sensitive situations, rather than doubt in the technology itself (Panda et al., 2019, p. 204).

Addressing these concerns, Buhmann (2023, p. 3) identifies three main problem areas in AI communication: evidence, outcome, and epistemic concerns. Evidence and outcome concerns relate to data quality and the fairness or accuracy of AI outputs, including risks of biased or incorrect results (Buhmann, 2023, p. 3). Epistemic concerns, in contrast, relate to limited transparency and traceability in AI systems, which make their decision-making difficult to understand or explain (Buhmann, 2023, p. 3). Hermann (2022, p. 1269) supports this view by describing many AI systems as “black boxes”, while Loof (2025, p. 181) adds that AI’s complexity and non-linearity can exceed human understanding. This lack of intelligibility raises ethical concerns about transparency and the justification of decisions (Hermann, 2022, p. 1269). In crisis communication, these issues are intensified by time pressure and incomplete and fast-changing information, which limit the ability to detect or correct AI-related errors (Prah & Goh, 2021, p. 5). As a result, AI can undermine organisational sense-making and credibility and turn technical failures into reputational crises (Prah & Goh, 2021, p. 5).

Bowen’s (2024, p. 8) empirical study further grounds these concerns in practice. The research shows that both communication practitioners and AI developers identify similar ethical risks, particularly regarding misinformation, black-box decision-making, privacy, and bias (Bowen, 2024, p. 8). These findings reflect Buhmann’s (2023, p. 3) evidence, outcome, and epistemic concerns, as participants describe how AI-driven misinformation disrupts organisational operations and creates security risks, while the lack of transparency in AI systems makes it hard to question or correct decisions in high-stakes situations (Bowen, 2024, p. 8).

Schneider and Boman (2025, p. 140) argue that the ethical application of AI in crisis communication relies on perceptions of authenticity, appropriateness, and crisis severity. Their case study on the Michigan State University shooting

demonstrates that generative AI during high-severity crises involving loss of life can trigger moral outrage, as stakeholders expect human empathy and authentic authorship (Schneider & Boman, 2025, p. 146). This implies that while AI might be acceptable in pre-crisis or low-severity situations, its use in sensitive contexts is often viewed as unethical and dehumanising (Schneider & Boman, 2025, p. 143). Piller (2025, p. 47) expands this view, showing that ethical judgements are influenced not only by message *content* but also by perceptions of the *creation process*. He analyses the same school shooting case, arguing that revealing AI use diminishes perceptions of “rhetorical humanity” as audiences perceive the message as lacking effort and care. Even if the language appears human-like, knowing it was generated by AI alters its interpretation (Piller, 2025, p. 47). Similarly, Kirk and Givi (2025, pp. 6, 8–9) suggest that emotional messages perceived as AI-generated are considered less authentic, cause moral discomfort, and provoke more negative reactions.

Ethical concerns about AI in crisis communication are linked to expectations that organisations visibly dedicate human time and emotional effort when addressing collective tragedies. Nies and Zhao (2025, p. 127) highlight that a key difference between human and AI communication is in empathy and personalisation. Human messages often use emotional intelligence to create connection and relatability, whereas AI content lacks true empathy and instead mimics emotion through pattern recognition rather than genuine understanding (Nies & Zhao, 2025, p. 127). Falkheimer and Heide’s (2023, p. 410) trust capital concept explains these responses by indicating that crisis messages are evaluated based on perceptions of an organisation’s credibility and caring nature. Consequently, the lack of visible human effort in AI communication might undermine trust, especially when audiences expect sincerity and empathy.

2.2.4. Human Versus AI-led Communication Tasks

Given this ethical context, it is essential to consider how communication tasks should be allocated between humans and AI systems. Since AI raises responsibility issues and practitioners may find it difficult to identify AI-generated content, it is crucial to distinguish which tasks AI can or should handle and which must remain human-led (Seidenglanz & Baier, 2023, p. 126). Johnson and Verdicchio (2017, p. 2268) define “AI anxiety” as the fear that AI might become uncontrollable or threaten human

uniqueness. They note that this fear often results from misunderstandings about AI autonomy and a neglect of the human, social, and institutional contexts surrounding AI development and use (Johnson & Verdicchio, 2017, p. 2268). Swiatek and Galloway (2022, p. 355) observe similar concerns in public relations, where automation is unsettling because the field relies heavily on human communication. This discomfort reflects wider concerns about interpersonal interactions, client trust, and the potential consequences of AI representing organisations (Swiatek & Galloway, 2022, p. 355). Additionally, fears about job losses and stakeholder treatment imply that increasing AI deployment could hinder relationship-building and trust (Swiatek & Galloway, 2022, p. 355).

Panda et al. (2019, p. 202) observe that practitioners are generally comfortable delegating routine tasks such as monitoring, sentiment analysis, updating media lists, and scheduling meetings to AI. AI is also appreciated for supporting campaign planning by detecting trends, analysing large datasets, profiling target audiences, and choosing suitable channels and influencers (Panda et al., 2019, p. 202). However, Vallor (2015, p. 108) cautions that heavy reliance on automated systems might lead to “moral deskilling”, where individuals lose the habit of actively considering the ethical aspects of decisions when technology handles parts of the judgement process. In communication practice, this raises concerns about whether practitioners may become overly dependent on AI-driven insights rather than critically assessing their broader impact on stakeholders (Sutherland, 2025, p. 64).

Nevertheless, practitioners tend to oppose AI more strongly during crisis situations, emphasising that while AI can identify emerging issues and analyse reactions, decisions on responses and messaging should be human-led (Panda et al., 2019, p. 204). Clients expect to see human involvement during crises, and culturally sensitive scenarios leave little space for automated judgement or mistakes (Panda et al., 2019, p. 204). Therefore, tasks related to trust-building and managing relationships with clients or stakeholders are consistently viewed as essential human responsibilities that AI can assist with but not replace (Panda et al., 2019, p. 209).

2.3. Responsibility and AI in Crisis Communication

The main challenge in assigning responsibility for AI-supported communication is that the process is no longer solely human-to-human. As Gunkel (2012, p. 19) points

out, traditional communication theory views communication messages as created by people and simply transmitted via technology, and it considers the computer a “neutral channel of message transfer”. However, this view is increasingly challenged as technology evolves beyond merely a “channel” and begins to actively contribute to interactions (Gunkel, 2012, p. 23). Since AI systems can now generate and influence messages independently, it becomes difficult to determine whether responsibility for the output rests entirely with humans or partly with AI (Gunkel, 2012, p. 23).

Before exploring how responsibility is managed in crisis communication, it is helpful to review discussions of AI responsibility in other high-stakes areas, for example autonomous vehicles. Cunneen et al. (2019, p. 721) point out that autonomous vehicles operate in morally charged contexts but lack moral reasoning capabilities, meaning responsibility for harm cannot be placed on the vehicle itself. Instead, it falls on the human and institutional frameworks involved in the design, programming, and deployment of these vehicles (Cunneen et al., 2019, p. 722). However, Hevelke and Nida-Rümelin (2015, p. 626) challenge this view, arguing that assigning responsibility to humans is justified only if they had a real chance to control or influence the system’s behaviour. Blaming individuals without such control is morally unjustifiable (Hevelke & Nida-Rümelin, 2015, p. 629). While they admit that choosing to use potentially risky technologies such as self-driving cars may entail some moral responsibility even without direct control (Hevelke & Nida-Rümelin, 2015, p. 628), they maintain that responsibility for autonomous systems is conditional, not automatic.

Furthermore, Schoenherr and Thomson (2024, p. 61) present an experimental study indicating that responsibility is not always clearly perceived or allocated in practice. Their two experiments involved asking participants to assign responsibility among human operators, developers, and AI systems in failure cases involving autonomous technologies, and the results showed frequent overlaps in responsibility across multiple actors (Schoenherr & Thomson, 2024, pp. 63, 66). This implies that, although AI does not eliminate human responsibility, autonomous systems add complexity to how responsibility is understood and attributed when failures occur (Schoenherr & Thomson, 2024, p. 69).

Hedlund and Persson (2025, p. 4056) argue that responsibility for AI cannot be understood only through intent or direct causation but must also be assessed regarding how different actors contribute to its development and use. Based on

interviews with experts in AI development, ethics, and policy, they identify contribution as central to how responsibility for AI is understood (Hedlund & Persson, 2025, p. 4056). Their respondents primarily linked responsibility to developers, who shape how AI systems function, but also stressed that users share responsibility because AI can learn from user input (Hedlund & Persson, 2025, p. 4056). This suggests that communication professionals who use AI tools during crises, for instance, to generate messages or input sensitive information, also share responsibility for the outcomes.

Zerfass et al. (2020, p. 384) emphasise that unclear responsibility structures pose a major challenge in AI-supported communication management, as professionals often struggle to assign responsibility for AI-related actions and decisions. This ambiguity can hinder decision-making and heighten ethical concerns within organisations (Zerfass et al., 2020, p. 385). The study also highlights issues regarding how employees and clients perceive AI in terms of trust and transparency (Zerfass et al., 2020, p. 385). Given that stakeholder trust and acceptance are vital for organisational legitimacy and adaptability (Zerfass & Viertmann, 2017, p. 75), negative perceptions of AI as untrustworthy or opaque can erode an organisation's relational capital and its ability to respond effectively to crises.

To address these concerns, the European Commission's Ethics Guidelines for Trustworthy AI² recommend maintaining human agency and oversight in AI operations, because this ensures that humans remain involved in decision-making processes (Bowen, 2024, p. 3; European Commission, 2019). While "human-in-the-loop" decisions may not be possible every time practitioners use AI, the guideline emphasises that moral judgement should not be delegated entirely to technological systems (Bowen, 2024, p. 3). Nevertheless, as the empirical studies above suggest, maintaining human oversight in practice remains challenging when responsibility structures are unclear and institutional frameworks are absent or incomplete.

² The guidelines identify seven requirements for trustworthy AI: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity and fairness, societal and environmental well-being, and accountability (European Commission, 2019).

2.4. Synthesis

This literature review shows that AI is already integrated into crisis communication as a supportive tool for monitoring and analysis. In strategic communication context, AI can be understood as a technological tool which can reason and analyse information in human-like ways and according to a specific purpose. While these functions can greatly enhance efficiency and situational awareness in a crisis, they also introduce ethical risks that are related to bias, transparency, misinformation, and trust. These issues are especially critical in crisis contexts. On one hand, current research indicates that practitioners are willing to rely on AI for routine tasks. On the other hand, studies also show that communication experts tend to resist AI, particularly in crisis responses, because these contexts require empathy and visible human involvement. Furthermore, previous research indicates that AI challenges traditional views of communication. If technology has previously acted just as a “channel” that helps to transmit messages, then AI actively influences how situations are interpreted and how messages are constructed. As a result, AI cannot be treated as a neutral tool, and because of this, responsibility becomes harder to locate, since AI itself lacks moral agency, but human responsibility depends on meaningful control and the ability to intervene.

3. Theory

This chapter presents the theoretical framework guiding the analysis of how Estonian communication professionals perceive ethical AI use and responsibility in crisis communication. The goal is not to judge whether the application of AI in crises is right or wrong, but to offer conceptual tools for understanding how responsibility is reasoned in practice. As AI increasingly influences communication by affecting how information is created and understood, traditional notions of ethics and responsibility become more complex. Therefore, the focus is on two key frameworks: 1) moral responsibility and its three subthemes, such as attributability, answerability, and accountability, and 2) machine ethics.

3.1. Moral Responsibility and its Three “Faces”

Before exploring moral responsibility in detail, it is helpful to understand the concept of responsibility itself. It is a layered term with various meanings depending on the context (Giddens, 1999, p. 8). Firstly, responsibility can refer to causal responsibility, meaning that an individual is responsible for an event because they caused or authored it (Giddens, 1999, p. 8). Secondly, it can denote ethical responsibility, which concerns whether someone has acted morally appropriately (Giddens, 1999, p. 8). Thirdly, and most crucially for Giddens (1999, p. 8), responsibility can be seen as obligation or liability, that is, being responsible for outcomes, especially in situations involving risk and uncertainty. Frankfurt (1969, p. 829) supports this view further by arguing that responsibility does not depend on the ability to have acted differently. What matters is whether an action reflects the agent’s own reasons or judgement, not whether alternative options were available (Frankfurt, 1969, p. 832). This idea is especially relevant in crises, where decisions are often made quickly and with limited options. The fact that professionals face constraints does not eliminate responsibility but shifts the focus to how decisions are explained and justified (Frankfurt, 1969, pp. 832–835). These various meanings demonstrate that responsibility encompasses multiple dimensions, including causal involvement, assigned obligations, and participation in

decision-making. However, not all forms of responsibility involve moral judgement, as causing something does not automatically mean someone should be blamed (Fischer & Ravizza, 1998, p. 2). To properly evaluate when a person should be judged morally, it is necessary to focus specifically on moral responsibility.

Moral responsibility is the basis on which a person may be praised or blamed for what they do (Fischer & Ravizza, 1998, p. 2). The focus is not only on *what* happened, but on *how* the person relates to their action: what matters is whether the decision reflects the person's judgement and reasons, not whether they had full control or could have acted differently (Fischer & Ravizza, 1998, p. 1). Moral responsibility should thus be distinguished from *causal* responsibility because an actor may causally contribute to an outcome without being morally responsible for it if they lacked the relevant understanding or ability to assess reasons during the action (Fischer & Ravizza, 1998, p. 2). Duff (2009, p. 978) complements this distinction by arguing that moral responsibility is often analysed in two main terms: control and knowledge. His examples illustrate that even if a person's actions causally contribute to harm, blame is not appropriate when the person lacked control over what happened or could not have known the relevant facts (Duff, 2009, p. 978). For example, a small child who *accidentally* knocks over a glass of milk is causally responsible for breaking the glass, but it would not be appropriate to blame the child in the same way as an adult who *intentionally* broke it (Fischer & Ravizza, 1998, p. 1). Moral responsibility, therefore, requires the agent to make conscious decisions and understand the significance of their actions (Fischer & Ravizza, 1998, p. 2).

According to Watson (1996, p. 229), moral responsibility has two main sub-themes, also known as "the faces of responsibility": attributability and accountability. Attributability is about whether an action shows something about a person's character and values, for example, an action is attributable when it reflects how a person thinks and what kind of person they are (Watson, 1996, p. 229). Accountability, on the other hand, concerns the social practice of holding people responsible for their actions, which involves making demands on them and responding to failures with blame or criticism (Watson, 1996, p. 230). Smith (2015, p. 109) agrees that attributability and accountability are different kinds of responsibility and further explains that attributability concerns judging what an action says about a person's character, while accountability asks whether it is fair to blame or criticise the person for that action.

However, Shoemaker (2011, p. 603) argues against Watson’s (1996, p. 229) dual distinction of moral responsibility and notes that only two forms do not capture all aspects of moral responsibility. Therefore, Shoemaker (2011, p. 603) introduces answerability as a third form. Answerability concerns whether it makes sense to ask someone to justify their actions by giving reasons (Shoemaker, 2011, p. 603; Talbert, 2024). Smith (2015, p. 103) also places strong emphasis on answerability and argues that an actor is answerable when it makes sense to ask them to explain and justify what happened, regardless of whether they exercised full control over the outcome. Thus, moral responsibility is useful for this thesis as it offers a multi-dimensional tool for examining how communication professionals themselves understand responsibility regarding AI in crisis communication. Rather than assuming a single meaning of responsibility, the distinction among attributability, answerability, and accountability enables to identify what kind of responsibility practitioners refer to when they describe AI-related decisions and failures, as shown in Figure 2.

Figure 2

Faces of Moral Responsibility

Responsibility	Core Focus	Example
Attributability	Whether an action can be meaningfully attributed to an actor	How a person interprets what an outcome reveals about professional judgement, competence, values or established ways of working
Answerability	Whether an actor can reasonably be asked to explain or justify an action	How a person identifies who is expected to give reasons , explanations, or accounts when outcomes are questioned
Accountability	Whether an actor can be held liable to blame	How a person discusses blame, fault or consequences when errors or harm occur

Note. Faces of moral responsibility. Adapted from “Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility”, by D. Shoemaker, 2011, *Ethics*, 121(3), pp. 603–604 (<https://doi.org/10.1086/659003>). Copyright 2011 by The University of Chicago Press.

Van de Poel (2011, pp. 38–39) similarly argues that responsibility is not a single concept but a cluster of meanings, some of which are descriptive (e.g., responsibility-as-task, responsibility-as-authority, responsibility-as-cause), while others are normative (they involve judgement, like blame, obligation, liability). This supports the need to separate different responsibility dimensions in the analysis, since practitioners may use the word “responsibility” to refer to different concepts depending on the situation.

For example, professionals may talk about what an AI-supported outcome reveals about their judgement or professional role (attributability), about who is expected to explain or justify the decision to use AI (answerability), or about who should be blamed or face consequences when something goes wrong (accountability).

Furthermore, when discussing responsibility, it is important to clarify the difference between moral and legal responsibility (Morse, 2004, p. 35). The distinction is especially relevant for accountability, since it involves blame and consequences and can therefore easily be confused with legal responsibility. In the framework used here, accountability follows Watson's (1996, p. 230) definition and refers to a moral and social practice of holding people responsible through blame, criticism, like reputational damage, loss of trust, or professional sanctions within an organisation. Thus, moral responsibility, more broadly, concerns whether a person deserves praise or blame for what happened, and this usually depends on whether the person had sufficient knowledge and control in the situation (Duff, 2009, p. 978). Legal responsibility, however, is formally determined and enforced by the state through laws and can result in consequences such as fines or imprisonment (Duff, 2009, p. 979). Duff (2009, p. 979) also argues that legal responsibility, especially in criminal law, is often narrower than moral responsibility because the law usually requires proof that a person acted voluntarily and with a certain level of fault. While people may still expect an explanation in everyday moral life, even when harm was accidental, legal systems usually do not require an account until it is first proven that a wrongful act happened and that the person meets the legal conditions for fault (Duff, 2009, p. 979).

3.2. Machine Ethics and Responsibility

Machine ethics refers to the study and development of machines designed to act in ethically acceptable ways, either by following human-defined rules or by applying ethical principles in decision-making (Anderson & Anderson, 2007, p. 15). The concept builds on earlier work in computer ethics, which focused on the social consequences of computing technologies and the ethical responsibilities of their human users (Moor, 1985, p. 266). Machine ethics extends this focus by shifting attention to the ethical significance of machine behaviour itself, especially as systems become more autonomous and operate with less direct human control (Anderson & Anderson, 2007, p. 15).

Many scholars argue that machines require ethical considerations because they are no longer merely tools. Hall (2011, p. 30) explains that machines need ethics since their actions are not always directly linked to human decisions. As systems grow more autonomous and complex, their behaviours can result in outcomes that humans cannot fully anticipate or manage (Hall, 2011, p. 42). He warns that unethically operated machines could cause significant harm, describing this risk as the danger of “superhuman sociopaths” (Hall, 2011, p. 42). Therefore, machine ethics is essential for guiding and limiting machine actions when human oversight alone may not sufficiently control them.

Moor (2011, p. 16) supports this argument by claiming that machine ethics has become necessary because computing technologies increasingly participate in decisions with ethical consequences. He describes computing technologies as inherently value-laden because they are designed to achieve specific goals and evaluated on whether they perform those tasks correctly (Moor, 2011, p. 14). Because machines perform tasks on behalf of humans, they can influence outcomes that matter ethically (Moor, 2011, p. 14). To clarify this role, Moor (2011, pp. 15–18) distinguishes four types of machine agents to separate ethical impact from moral responsibility: ethical-impact agents, implicit-ethical agents, explicit-ethical agents, and full-ethical agents. Moor (2011, p. 19) argues that simply having ethical effects does not make a system a moral agent. Many computing systems are involved in ethically significant situations, but they cannot reflect on or justify their actions (Moor, 2011, p. 15). By separating ethical impact from moral agency, Moor (2011, p. 18) shows that machines can be evaluated in ethical terms without being treated as responsible actors.

Based on these ideas, it can be argued that responsibility cannot be assigned to AI itself. When AI-supported crisis communication leads to harmful outcomes, responsibility must instead lie with human decision-making and/or organisational practices. Anderson and Anderson (2007, p. 19) further support this position by arguing that moral responsibility requires capacities such as intentionality and free will, which current AI systems lack. Simultaneously, they distinguish responsibility from ethical behaviour: while AI cannot be held morally responsible, it can still behave in ethically acceptable ways by following rules or principles (Anderson & Anderson, 2007, p. 19).

3.2.1. AI and Responsibility Gap

However, other authors show that this human-centred understanding of responsibility becomes difficult to apply in practice as AI systems grow more advanced. Sullins (2011, p. 152) argues that highly autonomous systems no longer function merely as tools but participate in situations with moral significance. He suggests that responsibility may appear to be distributed among designers, users, organisations, and systems, even though most moral responsibility still rests with humans (Sullins, 2011, p. 155). Johnson (2015, p. 708) further develops this argument by showing that autonomous systems make it difficult to trace their outputs to specific human decisions. Drawing on Matthias (2004, p. 181), she refers to this problem as a “responsibility gap”, meaning situations in which AI systems contribute to harm but no human actor can be clearly identified as responsible (Johnson, 2015, p. 708). Importantly, this gap does not arise because machines become moral agents, but because technological complexity undermines clear responsibility (Johnson, 2015, p. 713).

Santoni de Sio and Mecacci (2021, p. 1060) further clarify this problem by showing that responsibility can become unclear in several different ways when AI is involved. They identify a culpability gap in which harm occurs, but no individual can reasonably be blamed (Santoni de Sio & Mecacci, 2021, p. 1062). They also identify a moral accountability gap, where it is unclear who is expected to explain and justify what happened (Santoni de Sio & Mecacci, 2021, p. 1064). Additionally, they describe a liability gap, which concerns uncertainty about legal or organisational responsibility, and a role-responsibility gap, which arises when it is unclear who should supervise or intervene in the use of AI (Santoni de Sio & Mecacci, 2021, pp. 1065, 1067). To address these shortcomings, Santoni de Sio and Mecacci (2021, p. 1076) propose the concept of “meaningful human control”. This means that humans must be involved in the use of AI in a way that ensures both control over the system and a sufficient understanding of how it operates (Santoni de Sio & Mecacci, 2021, p. 1076). Their “tracing” condition requires that for every system action, at least one human agent can be identified who understands the system and can be held morally responsible (Santoni de Sio & Mecacci, 2021, p. 1077).

Nevertheless, Tigard (2021, p. 440) questions whether responsibility for AI must always be traced to a single person. Approaches like “meaningful human control” aim to address this by ensuring there is always identifiable human oversight, but

Tigard (2021, p. 441) argues that this relies on a too narrow understanding of responsibility and assumes that responsibility must come from individual agency and direct control. Instead, he suggests a more flexible approach: responsibility should not belong to just one person or a thing but be shared among many parties and seen as something that develops through practice (Tigard, 2021, p. 443). In this view, people connected to the AI, such as users or developers, can still be held responsible even if they do not have full control, because they consciously chose to use the system (Tigard, 2021, p. 443). Simultaneously, the AI itself can be part of how humans handle responsibility, for example, humans can restrict or retrain the systems after harmful outcomes (Tigard, 2021, pp. 442–443). While this does not completely solve the problem of unclear responsibility, it helps to manage it by spreading responsibility around and adjusting it based on each situation (Tigard, 2021, p. 444).

Thus, various authors indicate that questions of moral responsibility with AI remain quite ambiguous. This uncertainty can be understood through the concept of “anomie”, a sociological state in which weakened social norms leave individuals without clear behavioural boundaries or realistic limits on their desires (Marks, 1974, p. 333). In the context of AI, anomie reflects the idea that technological progress is outpacing the moral and institutional structures originally designed around human agency, leading to increasingly unclear rules concerning ethics and responsibility (Shad, 2025, p. 891). Moreover, AI’s rapid development has outpaced political and administrative processes intended to regulate it, creating a “field of tension” in which AI-driven measures might threaten the very democratic norms they aim to protect (Shad, 2025, p. 892).

4. Methodology

This chapter outlines the methodological framework of the thesis. It first presents the interpretive and social constructivist paradigm guiding the research, followed by the qualitative research design. The chapter then describes the process of collecting the empirical material, including the use of semi-structured, vignette-based interviews and the purposive selection of 16 Estonian communication practitioners. Next, it explains the abductive approach used to analyse the material and addresses the study's ethical considerations and reflexive practices.

4.1. Research Paradigm

Interpretive traditions, according to Prasad (2018, p. 14), are grounded in phenomenology and treat human meaning-making as the starting point for knowledge about the social world. They assume that social reality is constructed through the meanings people attach to objects or interactions and therefore prioritise understanding and human behaviour over causal explanation (Prasad, 2018, p. 14). This thesis departs from interpretive traditions because it prioritises understanding how practitioners themselves make sense and understand ethical use of AI and responsibility in AI-supported crisis communication. It assumes that the ideas about these dilemmas are shaped through meaning-making processes in professional communication practice and are not determined by causal or technical factors alone (Prasad, 2018, p. 14).

More specifically, this study draws on a social constructivist epistemology. Berger and Luckmann (1966, pp. 13, 78) argue that reality is socially constructed, and although the social world appears objective and independent, its objectivity is “a humanly produced, constructed objectivity”. Thus, what appears fixed and external has, in fact, been created by people – social phenomena do not exist on their own, separate from human beings, but are produced through human actions and maintained through ongoing interaction (Berger & Luckmann, 1966, p. 78). Similarly, Burr (2015, pp. 4–5) emphasises that what we regard as knowledge or truth is not derived from the world as it simply is but is constructed through everyday social interaction

and sustained through language and shared practices. To illustrate this, Berger and Luckmann (1966, p. 75) describe two individuals, A and B, whose repeated interaction leads to routines and shared expectations that gradually become taken for granted and appear stable. As they continue to act in similar ways, their behaviour becomes familiar and predictable, and the patterns they create together begin to seem normal (Berger & Luckmann, 1966, p. 74).

Building on this understanding, the ethical use of AI and responsibility in AI-supported crisis communication are viewed not as fixed or purely technical issues, but as socially constructed phenomena that evolve through communication practitioners' interactions and interpretations. Reflecting the social constructionist view of crisis communication by Falkheimer and Heide (2006, p. 180), crises are understood as sensemaking processes in which reality is negotiated within specific cultural and organisational contexts. Therefore, the use of AI is considered part of this larger framework, with its ethical use and responsibility shaped by how practitioners interpret and respond to particular crisis situations.

4.2. Research Design

According to Creswell (2007, p. 37), qualitative research is grounded in philosophical assumptions and seeks to understand the meanings individuals or groups attach to a social or human problem. This requires the researcher to approach the social world with the assumption that no detail is insignificant, and that even seemingly ordinary features of everyday practice may provide insight into deeper patterns of meaning (Bogdan & Biklen, 1998, p. 6). Rather than taking actions or statements for granted, the researcher continually asks why things are done in ways and what meanings participants attribute to them (Bogdan & Biklen, 1998, p. 6).

Therefore, the research design of this thesis adopts a qualitative approach. This choice reflects not only the use of in-depth interviews as a method but also a particular understanding of how research is conducted and how knowledge is generated (Flick, 2023, pp. 109–110). A qualitative approach is appropriate because the study aims to develop a detailed understanding of how ethical AI use and responsibility are interpreted and negotiated. As Creswell (2007, p. 40) argues, qualitative research is especially suitable when a phenomenon needs to be explored in detail through direct engagement with participants, allowing them to describe situations in

their own words rather than through predetermined categories derived from the literature. This is particularly relevant here, as responsibility in crisis communication is shaped by contextual conditions and organisational pressures that cannot be meaningfully separated from practitioners' own accounts (Creswell, 2007, p. 40). Additionally, current research does not provide a sufficiently detailed understanding of how communication professionals assign moral responsibility between humans and AI during crises, given that AI is still an evolving and rapidly advancing field of study (e.g., Hedlund & Persson, 2025; Schoenherr & Thomson, 2024). Considering this, an interpretive, meaning-centred approach offers a more appropriate strategy than quantitative measurement for capturing the complexity of these processes (Creswell, 2007, p. 40).

4.3. Collection of Empirical Material and Interview Proceedings

Empirical material was collected through in-depth semi-structured interviews, which used an interview guide with themes and key questions while still leaving room for the conversation to develop flexibly (Flick, 2023, p. 203). Semi-structured interviews are appropriate when the purpose is to gain an in-depth understanding of participants' experiences and perspectives (Tracy, 2020, p. 158). Instead of forcing respondents into fixed response categories, the interviews are designed to encourage them to explain their viewpoints in their own words, which makes it more likely that their own meanings and priorities will emerge (Flick, 2023, p. 203). Brinkmann and Kvale (2015, p. 27) add that a semi-structured interview seeks to understand the participant's lived, everyday world from their own perspective. It follows specific themes but remains flexible, allowing the conversation to develop naturally rather than following a fixed questionnaire (Brinkmann & Kvale, 2015, p. 27). Although it resembles an ordinary conversation, it is conducted with a clear professional purpose and is recorded and transcribed for later analysis (Brinkmann & Kvale, 2015, p. 27). Nevertheless, Alvesson and Kärreman (2011, p. 101) note that interview answers are always given in a specific social context, and the aspects influencing the interviewee cannot always be controlled. Therefore, it is important to keep in mind that participants are usually not stating objective facts, but their answers are shaped by their own identity, previous

experiences in the field, hidden motives and norms as well as the conversation between the interviewer and the interviewee (Alvesson & Kärreman, 2011, p. 102).

The interview guide used in this thesis followed three themes aligned with the research questions (Appendix 1). Firstly, participants discussed the use of AI in a fictional crisis scenario. Secondly, they argued about which tasks are best suited to AI and which must remain human-led. Lastly, they reflected on responsibility through a developed fictive crisis scenario. The interviews were held in Estonian and lasted 45–60 minutes each. As mentioned, throughout the interviews, a fictive but realistic crisis vignette³ was used as a thought exercise to prompt participants' reflections and help them apply their theoretical ideas to practice. It is important to note that the vignette itself is not the basis of the analysis. Rather, the analysis is structured around themes that emerged from participants' reasoning and are illustrated by the thoughts shared during the discussions of the vignette. The vignette served solely as a stimulus. Jenkins et al. (2010, p. 178) argue that the purpose of qualitative vignette interviewing is not to predict what an interviewee would do, but to understand the perceptions through which they make sense of a situation. By asking participants to place themselves in the role of a communication professional in a crisis, the vignette encouraged them to imagine the situation from within and to reason through it as they would in practice. This makes it possible to access reflections and assumptions about professional responsibility and AI use that participants may not have shared without a concrete scenario (Jenkins et al., 2010, p. 181).

4.3.1. Selection of Participants

Firstly, participants were selected using purposive sampling combined with criterion sampling. According to Patton (1990, p. 169), purposive sampling involves deliberately selecting information-rich cases that can provide deep insight into the phenomenon. The focus was on identifying practitioners with relevant experience on the research topic (Patton, 1990, p. 169). In addition, criterion sampling was used, whereby participants were selected based on predefined inclusion criteria (Patton, 1990, p. 176). The key criteria were practical experience in crisis communication and the use

³ The vignette draws on a 2017 Estonian ID-card security risk affecting approximately 750,000 users (Vahtla, 2017). In Estonia's digital society, such incidents threaten trust and reputation, making the scenario a realistic test of how communication professionals understand AI use and assign responsibility under pressure.

of AI tools in professional communication. Only practitioners who met both conditions were included, as they were considered most likely to provide empirically grounded, information-rich insights. Practitioners were asked prior to the interviews to confirm their experience in crisis communication and use of AI. This ensured that all participants could draw on concrete professional situations rather than hypothetical reflections (Patton, 1990, p. 170).

Given the small size of the Estonian communication field, access to relevant participants was partly facilitated through professional networks. After identifying and interviewing the first participants who met the inclusion criteria, snowball sampling was also applied. Patton (1990, p. 176) describes snowball sampling as a strategy in which existing participants recommend other individuals who can contribute valuable perspectives. In a small, connected professional community like Estonia, this approach was appropriate because practitioners with relevant experience in AI and crisis communication are often well known within professional circles. The selection process aimed to capture variation in professional experience across junior, middle, and senior communication professionals, as shown in Table 1.

Table 1
Participants, Experience and Length of Interviews

ID	Position	Experience	Length (minutes)
1	Middle level	6–10 years	54
2	Junior level	Up to 5 years	48
3	Senior level	Over 10 years	53
4	Junior level	Up to 5 years	55
5	Junior level	Up to 5 years	68
6	Senior level	Over 10 years	58
7	Senior level	Over 10 years	51
8	Senior level	Over 10 years	71
9	Middle level	6–10 years	49
10	Middle level	6–10 years	56
11	Senior level	Over 10 years	57
12	Middle level	6–10 years	46
13	Junior level	Up to 5 years	49
14	Senior level	Over 10 years	59
15	Middle level	6–10 years	46
16	Middle level	6–10 years	66

There is no unified framework for assessing the seniority of strategic communication practitioners based on years of experience. Previous public relations studies have, however, mostly distinguished among three levels of practitioners based on experience and strategic thinking. For example, Steyn (2018, pp. 2–3) identifies the roles of the technician, manager, and strategist, which provide a useful basis for distinguishing participants between junior, middle, and senior levels. Accordingly, seniority in this thesis was assessed by participants' years of experience in the field of communication. Participants with up to five years of experience were categorised as junior-level practitioners, those with six to ten years of experience as middle-level practitioners, and experts with more than ten years of experience as senior-level practitioners. Including practitioners from various levels was crucial because perceptions of ethics and responsibility in crisis communication can vary based on organisational role and experience. Thus, the selection included 16 Estonian communication practitioners from 13 public and private organisations, with four junior practitioners, six middle-level practitioners, and six senior experts.

As Brinkmann and Kvale (2015, p. 113) argue, the number of participants in qualitative interview research depends on what is needed to generate knowledge. Thus, the interviews should continue until saturation is reached (Brinkmann & Kvale, 2015, p. 113). They also note that interview studies often include around 15 participants, as larger samples may reduce the time available for in-depth analysis, while smaller ones may limit the breadth of perspectives (Brinkmann & Kvale, 2015, p. 113). Based on this, 16 interviews were conducted with Estonian communication professionals, as this amount provides sufficient variation in experience while still allowing for detailed interpretation of the empirical material.

4.4. Analysis of Empirical Material

Rather than following a linear sequence in which theory is developed first and then tested against empirical material, the analysis followed an abductive logic. Abduction is a process that begins when existing theoretical understandings do not fully explain empirical observations (Alvesson & Kärreman, 2011, p. 58). Thus, the study did not rely solely on patterns emerging from the material (inductive), nor did it begin with fixed ideas to be tested (deductive), but instead, analysis developed through ongoing dialogue between interview data and theory to better understand how ethical AI use

and responsibility are shaped in AI-supported crisis communication (Tracy, 2020, p. 27). As Eisenhardt et al. (2016, p. 1119) argue, the challenge for qualitative scholars is to remain systematic and transparent while avoiding methodological procedures so rigid that they limit insight. Overly mechanical application of codes may therefore produce studies that appear methodologically rigorous but are uninformative and contribute little theoretically (Eisenhardt et al., 2016, p. 1119). This is why qualitative analysis should remain flexible and iterative so that researchers can identify surprising and fruitful patterns in the material and develop interpretations that extend theoretical understanding.

According to Alvesson and Kärreman (2011, pp. 58–59), the abductive approach involves three steps: applying an established theoretical perspective, identifying empirical observations that appear surprising or insufficiently explained considering that perspective, and refining or reworking the interpretation to better account for the material. These three steps also structured the analytical process in this study. Before beginning the analysis, all interviews were transcribed, translated into English, and sanitised to remove any confidential information. The transcripts were then read several times to gain an overall understanding of the material. The analysis began by developing a coding frame to address the research questions, combining descriptive codes emerging from the interview material with theoretical concepts related to moral responsibility, machine ethics, and AI. Secondly, the empirical material was analysed iteratively using this coding frame. Attention was on how practitioners described the role of AI in crisis communication, how they justified assigning certain tasks to AI and how they understood responsibility when AI-related errors occurred.

4.5. Ethical Considerations and Reflection

Alvesson (2023, p. 89) emphasises that researchers should remain reflexive throughout the study and consider multiple interpretations rather than privileging a single reading of the material. Interview insights should therefore not be treated as straightforward representations of reality but should be interpreted carefully, as situational dynamics shape what is said (Alvesson, 2023, p. 97; Alvesson & Kärreman, 2011, p. 102). Ethical considerations were also central, as qualitative research affects other people and requires attention to permission, confidentiality, participation, researcher-participant relationships, and transparency (Tracy, 2020, p. 87). Before the interviews,

participants were informed of the study's purpose and their role in it, as well as their right to withdraw at any time. They received the information prior to the interview (Appendix 2).

As the Estonian communication field is relatively small, it was important to protect participants' identities. Therefore, all interviews were sanitised to remove any identifying or delicate information. In addition, pseudonyms were used. Raw interview material was stored securely and was accessible only to the researcher, and some contextual details were slightly modified in the final transcripts (Tracy, 2020, p. 89). Although full anonymity cannot be guaranteed in a small professional community, as some thoughts or sentences might reveal participants' identities, confidentiality has been ensured. This means that participants cannot be directly linked to the thesis through publicly available information.

Lastly, it is important to acknowledge that the researcher is the primary research instrument in a qualitative study (Tracy, 2020, p. 88). As the thesis examines ethics and responsibility in AI use, which is a topic often associated with potential judgement and critique, it was important to avoid framing questions in ways that could imply negative assessment (Alvesson, 2023, p. 116). Instead, neutral wording was used to encourage reflection and to maintain a respectful and supportive interview atmosphere.

5. Analysis

The following chapter analyses how Estonian communication practitioners understand and negotiate the questions of ethical AI use and responsibility. It is structured around the three research questions, with each theme illustrated by interview quotes from discussing the fictive ID card crisis vignette. Firstly, AI use in crisis communication is examined through four main themes that emerged from the interviews: routine work, AI for sensemaking, the “black box” problem, and AI governance. Secondly, the division of tasks between humans and AI is analysed, with a focus on crisis severity, verification capability, and data sensitivity. Lastly, questions of responsibility regarding AI use are explored through the framework of moral responsibility and its “three faces”.

5.1. Use of AI in Crisis Communication

This first chapter explores how practitioners use AI in crisis communication and how they think about it from an ethical standpoint. The interviews show that practitioners see AI as a task-specific support tool, not an all-purpose solution, while keeping strategic decisions and final judgement human-led. Four main themes emerged from the interviews: 1) the routine work, 2) AI for sensemaking, 3) the “black box” problem, and 4) governing AI through absent guidelines.

5.1.1. *The Routine Work with AI: Efficiency Versus Loss of Skills*

AI is described as useful for monitoring the communication environment and for supporting routine tasks during a crisis. These activities include analysing media coverage, tracking discussions across channels, and preparing materials under time pressure. Previous research identifies similar functions as key areas where AI can support crisis communication, particularly in monitoring stakeholder discussions and assisting with message production (Cheng et al., 2024, p. 177; Hancock et al., 2020, p. 91).

While practitioners in this study report using AI in comparable ways, what is particularly interesting is how they categorise these routine tasks:

“Honestly speaking I use AI more in case of such “meaningless” things – for example product communication or content-marketing, which is necessary quickly to get live and where risk-level is low.”

The characterisation of certain tasks as “meaningless” or less important reveals a professional hierarchy among routine tasks: some work is considered strategically significant and worth human effort, while other tasks are dismissed as routine production that can be delegated to AI. From an efficiency perspective, this is a rational choice: why spend professional time on routine drafting when AI can produce a usable first version in seconds. AI will increasingly become a tool for eliminating what experts consider low-value work: repetitive and time-consuming tasks that are not expected to require expert judgement (Cheng et al., 2024, p. 177). Yet several participants also acknowledged a troubling side effect of this delegation: the loss of professional skills and critical thinking. The following thought stood out:

“I would like to create some subtext or write something from scratch myself, but I probably do not do it anymore. I also have a fear of the blank page, and sometimes I feel that I can no longer think outside the box if I do not have chat [AI] to bounce with all the time.”

This example reveals a paradox of AI use. On one hand, AI is adopted to overcome creative blockages (e.g., “fear of the blank page”). However, it is precisely the use of AI that seems to create these same obstacles. The practitioner acknowledges that they no longer write from scratch most of the time, suggesting that the skill itself is disappearing. This indicates that AI use is not simply a neutral efficiency gain but a practice that reshapes professional capacities over time, and not necessarily for the better. This reflects what Vallor (2015, p. 108) describes as “moral deskilling”: a process in which, due to reliance on automated systems, the habit of actively reflecting on the ethical consequences of decisions gradually disappears as technology takes over part of the decision-making process. Thus, in crises, specialists risk losing not only technical skills, such as drafting texts from scratch, but also the ability to critically assess what AI-generated results mean for stakeholders and how they might escalate the situation (Sutherland, 2025, p. 64).

In addition to loss of critical thinking and skills, several practitioners also described that excessive AI use can lead to “mental laziness”. They noted that routinely relying on AI-generated text without critical review reduces the overall quality of communication work to the point that it becomes publicly noticeable. The language may appear weaker and less convincing, thereby reducing the message’s credibility. A practitioner said that poor-quality AI output is often detectable, as audiences and colleagues may recognise when content has been generated by AI. Another practitioner added:

“If you have a colleague who does not dive [in] deeply, then mental laziness arises. Empty sentences come [from AI] that mean nothing. You ask: Did you, yourself, read what this means? And if one does not dive in, then we do not have full certainty that it [AI] does not fantasise something in between.”

This fear of laziness shining through also reflects how stakeholders may perceive AI-generated communication. Delegating communication tasks to AI can create additional reputational risks, even when the tool’s response appears technically sufficient. Schneider and Boman (2025, p. 146) demonstrate that the use of generative AI in high-severity crises can indeed trigger moral outrage, as stakeholders expect human empathy and genuine authorship rather than efficiency-driven automation. Research by Kirk and Givi (2025, pp. 6, 8–9) also confirms this pattern: when stakeholders believe an emotional message was written by AI, they see it as less authentic and respond more negatively to the sender. Thus, this concern about “mental laziness” becomes even more important in crisis situations, where stakeholders are more sensitive to how messages are written and who stands behind them (discussed further in Chapter 5.2.1 regarding AI and crisis severity).

5.1.2. AI as Sensemaking Support, but Not a Sensemaker

Alongside routine work support, communication professionals described AI as useful in the early phase of a crisis, when the main task is not yet drafting public messages but understanding what is happening and deciding how the organisation should respond. At this stage, practitioners must interpret incomplete information, identify potential risks, and anticipate how stakeholders may react. AI is therefore used mainly to support this analytical work rather than to produce communication itself. In this role, AI helps practitioners explore different ways of interpreting the situation (Cheng

et al., 2024, p. 177). One practitioner described how AI can be useful in the fictive ID card crisis vignette:

“I would [...] ask it [AI] to identify risks [...] and what mitigation steps we should take. [...] To suggest possible guidance for people and if similar cases have occurred globally. [...] I would use it to build context and perhaps suggest a communication strategy. I would not rely on it 100%, but I would be interested in the angle it proposes.”

These are activities that align with what Weick (1988, p. 305) describes as sensemaking in crises, where incomplete and constantly changing information forces practitioners to form plausible explanations while events are still unfolding. AI is well-suited for these tasks because it can process large volumes of information and generate multiple interpretive frameworks quickly (Hao et al., 2025, p. 2). Nevertheless, there are also shortcomings in using AI for sensemaking. Although practitioners repeatedly emphasised that AI helps gather and make sense of the situation, they also noted that these tools are not limited to passive assistance. AI can actively influence which interpretations become public and credible. When practitioners ask AI to “identify risks” or “recommend strategies”, the result does not reflect neutral information but actively produces interpretations shaped by the system’s training data and algorithmic priorities. AI may highlight aspects that practitioners never intended to emphasise, subordinate important considerations, or rank priorities in ways that differ from experts’ understandings. This tension was acknowledged as follows:

“I have also used it for explaining legislative topics. For example, about a law being discussed in parliament, to get a first-level overview. But here is an important nuance – when I later spoke with my policy partners, it turned out that the emphasis points were in completely different places than what came out of the AI answer. If I had relied only on AI, I might have gone in the wrong direction.”

This reveals an interesting problem: practitioners might not often understand whether AI has led them astray until they check with other human sources. Instead of passively supporting human interpretation, AI creates its own meaning – one that may be systematically distorted yet appear authoritative precisely because practitioners cannot track how AI made these interpretive choices (Buhmann, 2023, p. 3). The practitioner only realised that AI had emphasised the wrong aspects after consulting with policy experts. Had this verification step not occurred, the AI-generated interpretation could have misinformed the practitioner’s understanding of the situation. This suggests that

AI does not support human sensemaking in a neutral way but, as Guzman and Lewis (2020, pp. 71–72) argue, it can function as a communicative agent that actively participates in meaning-making processes. Thus, when AI identifies crisis risks or suggests strategies, it is not merely helping practitioners see what is already there but is actively constructing a version of the situation that may differ from real-world understanding. The practitioner must then decide which interpretation to trust, but may often lack the expertise to make that judgement reliably. This awareness that AI may construct misleading interpretations leads practitioners to emphasise the need for critical evaluation of AI results against organisational knowledge:

“You still must think yourself whether it fits your organisation and the situation. It might give you ideas, but you cannot just take it as it is.”

This concern reflects the idea that Johnson and Verdicchio (2017, p. 2268) identify as one of the core limits of AI tools: they still operate outside the human, social, and institutional contexts in which crises occur. This concern about “fit” points to a practical limitation in how AI tools actually work – they generate suggestions based on the data they were trained on, but practitioners cannot easily see which sources the AI used or whether that information is accurate (Sutherland, 2025, p. 63). In fact, generative AI tools sometimes fabricate information or hallucinate (Sutherland, 2025, p. 63). This makes it difficult to determine whether an AI output is factually correct, let alone whether it makes sense for a specific organisation or a situation. As a result, AI cannot reliably propose answers that align with the organisation’s values, key stakeholders, or the cultural dynamics of a particular crisis. Practitioners must control and add this contextual understanding themselves.

5.1.3. The “Black Box” Problem

The problem with fabricated information and hallucinated answers leads to one of the main limitations of today’s AI tools regarding the reliability and transparency of AI-generated information. Despite describing AI as useful for routine work and sense-making with certain flaws, practitioners across the interviews consistently emphasised that they never rely on AI outputs without first checking the facts. This verification practice was described not as optional but as fundamental to responsible AI use. One participant explained that AI can never be entirely trusted, as it may hallucinate and

produce fictional answers. The idea of “trust but verify” appeared repeatedly in interviews. This distrust appears to stem from practitioners’ practical experience that AI systems behave in ways humans cannot understand, leading to uncertainty about how they produce certain outputs. Several participants described AI tools as “black boxes” – systems whose reasoning processes remain non-transparent and confusing to users (Hermann, 2022, p. 1269). A thought stood out:

“We have [...] such AI systems, in case of which I no longer understand why it does something. There exist tools where the “black box” is even bigger and unpredictability is significantly bigger. Even those mass-use tools are black boxes. I never know 100% why I receive exactly such a result.”

This lack of transparency poses a challenge about how to reason for and explain AI’s answers and how to determine responsibility for the output. For it to be fair to hold someone responsible, they must have known the relevant facts and have genuine control over what happened (Duff, 2009, p. 978). Yet when practitioners use these AI tools in crisis communication, these conditions are not met. They cannot know why AI produces certain outputs because the system operates as a non-transparent “black box” (Hermann, 2022, p. 1269). Neither can they fully control what AI generates because, as one practitioner emphasised, it may “add something with its own head”. Loof (2025, p. 181) further explains why this is the case: AI models, especially deep learning models, are described as “black boxes” because their functioning is so complex and nonlinear that it exceeds human capacity to fully understand it. Thus, practitioners may well understand the result produced by AI, it may even correspond to what they know to be true, but they cannot track how the tool arrived at that very specific conclusion. Loof (2025, p. 181) warns that this uncertainty raises serious ethical questions when AI is used to make critical decisions. In crisis communication, where a single misjudgement can escalate a crisis rather than contain it, practitioners are caught in a difficult position. They are held responsible for the communication they produce, yet they are using tools whose decision-making processes they cannot trace or explain.

5.1.4. Governing AI Through (Absent) Guidelines

The analysis so far has revealed that AI use in crisis communication, despite its convenience and efficiency, involves significant challenges: AI can hallucinate, erode

skills and critical thinking, and operate as a mysterious “black box”. Given these tensions, one might expect organisations to provide clear guidelines governing when and how AI should be used in crisis communication. However, the interviews revealed that organisations either lack formal guidelines entirely or have policies that remain incomplete and unenforced. While some participants work in organisations with data protection rules that indirectly constrain what information can be entered into AI systems, few reported having explicit policies on how AI should be used ethically.

Furthermore, the interviews revealed variation in how organisations approach AI adoption and governance. Participants working in private organisations reported that the workplace generally encourages AI adoption, framing it as essential for efficiency and competitiveness. One practitioner described how their management actively promotes the use of AI, arguing that it helps the organisation stay modern and frees employees’ time for higher-value work. However, public organisations operate under stricter data protection requirements and greater public responsibility, making them more cautious about using AI. One public sector participant explained:

“We have clear institution-wide rules about what can and cannot be entered, especially classified information. [...] We have not separately mapped out detailed AI risks. But since we already work strictly by the book and deal daily with confidential information, there is no temptation to put much into AI.”

Most of the participants’ organisations fall somewhere between active encouragement and security-focused restriction, having begun policy development that remains incomplete. The principle of avoiding sensitive data persists across organisations, yet it remains open to interpretation to practitioners: what counts as sensitive, and how sensitive is too sensitive. In addition, some experts have taken the initiative and developed guidelines on their own. One interviewee explained that they created a guideline by themselves for AI use inside the whole organisation:

“[...] We had some terms of use, but I added from my side that if you use it, remember certain things. For example, do not feed restricted documents into it. I also emphasised that you are always the final person who decides what goes out. I deliberately brought in this human factor.”

Without institutional guidance, practitioners are left to navigate AI ethics on their own. Some draw boundaries between public and internal content, others develop personal rules about what feels acceptable. Nevertheless, these individual approaches

only mask a deeper problem. One practitioner captured this precisely, saying that if AI guidelines are not clear and established at the organisational level, then one cannot blame only the employee if something goes wrong when using AI. This creates what Johnson (2015, p. 708) describes as a responsibility gap: a situation in which it becomes unclear who should explain failures, who should supervise ethical AI use, and whether users can reasonably be blamed (Santoni de Sio & Mecacci, 2021, pp. 1062–1067). When organisations fail to specify expectations and provide frameworks, they create conditions in which practitioners face blame for misjudgements for which they had no guidance on how to avoid. Therefore, the absence of guidelines does not leave a neutral space that practitioners fill with professional judgement. Rather, this creates a conflict of responsibility, in which individuals bear the responsibility for decisions that should have been regulated at the institutional level. Practitioners become solely responsible for navigating ethical dilemmas that top management has failed to address.

5.2. Task Division Between Human and AI

While practitioners agree that humans must have the final authority in all communication tasks, the actual division of tasks between humans and AI remains shaped by factors that extend beyond this principle alone. The following subchapter analyses how practitioners navigate this division of tasks. Rather than listing which specific tasks go where, the analysis explores the decision-making logic practitioners employ when deciding task appropriateness. The interview material reveals that practitioners do not apply a single universal rule but navigate task allocation through three main arguments: 1) crisis severity, 2) verification ability, and 3) data sensitivity.

5.2.1. Crisis Severity as Decision Threshold

When practitioners describe how they decide to use AI in a crisis, they consistently point to crisis severity as one of the most determining factors. One interviewee differentiated between genuine crises and less critical situations as follows:

“If it is a “real crisis”, then I always know what to say and act quickly. If it is a softer [...] or crisis-like situation, then it may be harder to find the right angle and there AI could help. But in a serious crisis, the focus is so clear.”

What practitioners see as a “real crisis” reflects Coombs’ (2006, p. 243) understanding of crisis severity: situations where the potential for physical, emotional, or environmental harm is substantial, or where the organisation’s reputation and financial stability face significant threat. Hence, when lives or organisational survival may be at stake, practitioners cannot afford to appear uncertain or dependent on AI-assistance. The decision to exclude AI is therefore not about technical capability but about maintaining professional credibility under conditions where, as Panda et al. (2019, p. 204) found, clients expect visible human involvement. Thus, sensitive situations leave little room for automated judgement or error that AI might create.

What makes this severity-based reasoning interesting is that it shows the opposite of the usual time-saving or efficiency benefits often used to justify AI (as discussed previously in Chapter 5.1.1). Previously, practitioners described delegating “meaningless” routine tasks to AI precisely because it saves time, freeing them for higher-value work. In fact, speed is AI’s primary appeal in crisis communication contexts (Hossain et al., 2025, p. 157). Yet when the severity of the crisis escalates, participants tend to consciously reject this advantage. Several explained that when the stakes are high, they would rather take an extra 10–15 minutes to verify information manually than risk making a fatal mistake. AI remains just as fast, but speed is no longer the priority. As one practitioner put it, in serious situations, a small error does more damage than taking extra time to verify. This shows that crisis severity serves as a threshold at which professional priorities fundamentally shift: the time-saving that makes AI valuable in routine work becomes irrelevant (or even dangerous) in serious crises, where accuracy and human touch matter more than speed.

Although time constraints raised concerns that the work might appear careless, the participants also discussed the direct consequences of reputational damage if stakeholders discovered that AI had been used. Earlier, practitioners worried that “mental laziness” would become detectable when they delegate too much routine work to AI, but when discussing tragic crisis situations, the concern deepened. Regarding the fictional ID card crisis vignette, a practitioner pointed out that if stakeholders discovered AI had been used in the communication process, they would see it as evidence that the organisation did not care enough:

“In crisis situations, one should not use it [AI] too eagerly. If there is truly a tragic situation, people have [...] suffered greatly, and then it comes out that you did not even bother to write the text yourself, but let a machine do it... It looks very bad.”

This reasoning aligns with Schneider and Boman's (2025, p. 143) claim that AI may be ethically acceptable in low-severity situations but is perceived as unethical in sensitive cases. What this thesis's practitioners add is how they anticipate this judgement to draw boundaries: severity determines not only whether AI can produce appropriate content, but also whether using AI in a situation can be publicly defended. This fear is also supported by Piller (2025, p. 47), who argues that the use of generative AI undermines perceptions of "rhetorical humanity" as audiences interpret AI involvement as a lack of personal effort and care. One can argue that this, in turn, can lead to a new reputational crisis because the issue is no longer only about the situation itself, but also about *how* the organisation responds to it.

5.2.2. *Verification Ability*

Additionally, the interviews revealed that the "black box" problem discussed in Chapter 5.1.3 not only shapes practitioners' general distrust of AI, but also directly determines which specific tasks they are willing to delegate to AI. Thus, in addition to crisis severity, practitioners assess whether they feel capable of sufficiently controlling AI's outcomes and use this as the primary criterion for deciding which tasks must remain under human control. This, once again, reflects what Duff (2009, p. 978) describes as the two basic conditions of moral responsibility: control and knowledge. Duff (2009, p. 978) argues that it is only fair to hold someone responsible if they are aware of the relevant circumstances and have actual control over what is happening. When practitioners discuss which tasks they do not delegate to AI, they are describing the boundaries within which they feel they can maintain that control and knowledge. In other words, the more complex and extensive the tasks become, the less willing practitioners are to hand them over to AI:

"[...] It is a question of control. If I have not written and thought it through myself, then I do not fully own it. [...] If I let AI produce it, I lose control. [...] I do not want the final text to be something where I do not know exactly how I arrived at it."

The idea suggests that "control" in this context requires truly thinking through the logic of the answer yourself, not just checking what the AI produced. Without this mental work, practitioners feel they do not "fully own" the communication. The worry about answering competently can be further connected to answerability – the ability to give reasons for decisions (Shoemaker, 2011, p. 603; Smith, 2015, p. 103). If AI

produced a communication strategy for a crisis and the practitioner did not work through it themselves, they may not be able to explain why a particular argument or recommendation was included. Thus, practitioners may use AI to draft, word, or structure text, but they still work through the content's logic themselves. This ensures that, if challenged, they can explain why the communication takes a particular position.

Simultaneously, the issue of knowledge creates additional challenges. Practitioners often tend to lack the technical expertise needed to fully verify AI-generated content on specific topics. Reflecting on a mistake described in the vignette, one practitioner explained:

“A communication person does not have to understand everything. [...] Over-checking is important both when you use AI and when you do not use it – because you do not know enough about that topic.”

This suggests that verification is not always the responsibility of an individual but often requires collaboration. Simultaneously, this blurs the line of moral responsibility. Although practitioners bear causal responsibility for the information they create (they did it), they may lack the specialised knowledge necessary to assess the accuracy of the content (they understood it) (Giddens, 1999, p. 8). In other words, they fulfil the condition of having caused the action, but not necessarily the condition of fully understanding it (Duff, 2009, p. 978; Fischer & Ravizza, 1998, p. 2). However, some participants reported addressing this issue by establishing additional levels of review. Material generated by AI is typically reviewed not only by communication specialists but also by subject matter experts within the organisation. This joint review process helps ensure that the final message is both accurate and well-reasoned, even if individual practitioners lack all the necessary knowledge.

5.2.3. Data Sensitivity

A third issue concerning the boundaries between AI and human tasks is what kind of crisis information can be fed into AI. A clear principle emerged during the interviews: only information that is already public or will soon become public may be used in AI tools. Regarding crisis situations, one practitioner noted:

“I share with AI data that I am ready to share with the public. If I have unconfirmed sensitive information that should not reach the public, then I would not put it there.”

This principle acts as a safeguard against responsibility-related consequences. If sensitive crisis information is leaked because it was inserted into AI, practitioners would be held responsible for the breach. As Hedlund and Persson (2025, p. 4056) similarly argue, liability for AI can be based on contribution: those who decide to input data into the systems influence the results and therefore share responsibility for the consequences. By limiting AI use to publicly available information, practitioners ensure that, even in the event of a data leak, their decisions would not result in further breaches of confidentiality.

However, interviews show that the boundaries of data sensitivity vary greatly depending on the organisational context and existing infrastructure. Private-sector specialists with access to secure, closed AI systems described being able to work with sensitive crisis data in ways that would otherwise be impossible. Several participants explained that their organisation has developed a fully internal AI environment, a “sandboxed” system that allows them to use any type of data without fear of it leaking. For those with the resources to build such infrastructure, data sensitivity becomes less of a technical constraint, and specialists can insert confidential details without the risk of external disclosure, meaning they can also use AI tools for more tasks in crisis communication. Public sector experts, however, mostly do not have access to “sandboxed” AI and must use public ones instead:

“We first have to say that the public sector [...] is too slow. Especially in the security field: highly regulated and cautious. That does not mean we do not want innovation, but currently there is a clear rule that classified information cannot be put into any [public] AI system.”

As a result, public sector employees face several restrictions. First, they often work with information that is legally classified for internal use only, such as details of security incidents or citizens’ personal data, which cannot be processed by public AI tools. Second, they often lack the organisational resources to develop “sandboxed” infrastructure that would enable the use of AI with sensitive data. The claim that the public sector is too slow also reflects an awareness of a structurally unfavourable situation: strict regulations clash with limited budgets and create conditions in which innovation is desired yet blocked by institutional barriers.

Nevertheless, one should not assume that a “sandbox” solution programmed specifically for a company will be problem-free. When practitioners use tools they believe are secure, they do so on the basis of their organisation’s assurances. But this

creates a knowledge limitation (Duff, 2009, p. 978) because practitioners cannot independently assess whether these systems are truly secure – they simply must trust what they have been told. For example, if sensitive crisis data leaks from the “sandboxed” tool despite the organisation’s security assurances, responsibility for the breach becomes unclear. The practitioner entered the data into AI precisely because management had assured the security of the infrastructure and encouraged its adoption. This once again illustrates the gap in moral responsibility: when a data breach occurs, it appears to be partly the result of multiple parties’ actions, yet no specific individual can be held fully responsible (Santoni de Sio & Mecacci, 2021, p. 1064). Responsibility may lie with the practitioner who entered the data, the management that assured employees it was safe and encouraged the use of AI, and the technical team that built the system. However, when responsibility is distributed across multiple people and systems in this way, assigning it fairly becomes extremely difficult (as discussed further in Chapter 5.3.3).

5.3. Understanding Responsibility with AI-related Errors

Having examined how practitioners establish boundaries for AI use and briefly touched on the uncertainty surrounding questions of responsibility, this chapter specifically explores how the interviewees perceive moral responsibility when AI-assisted crisis communication leads to problematic outcomes. Rather than assuming moral responsibility has a single meaning, the analysis employs the three-part framework introduced in theory to examine how practitioners specifically understand it through attributability, answerability, and accountability.

5.3.1. Attributability: Crisis Softens the Judgement of Actions

When participants were asked whether using AI inappropriately during a crisis indicates a colleague’s incompetence, practitioners consistently rejected this idea. They argued that one cannot draw such strong conclusions about someone’s professional character without knowing the circumstances they were working under. According to Watson (1996, p. 229), attributability refers to whether an action provides insight into a person’s professionalism and values. Practitioners, however, noted that decisions made in crisis situations should not be interpreted as evidence of stable professional

competence because in high-pressure moments, people do not behave as they normally would, and mistakes are more likely to happen:

“To assess their general professionalism based on that [mistake with AI] would be too strong conclusion, because we do not know in which conditions that decision was born. Maybe they had an exceptionally difficult day or something heavy had recently happened that influenced their attention.”

This reasoning goes hand in hand with Fischer and Ravizza’s (1998, p. 2) argument that moral responsibility depends on actions that reflect a person’s actual reasoning. In a crisis, situational constraints distort that reasoning, which is why errors made under such pressure should not be taken as a reflection of who the practitioner truly is. Several interviewees further reinforced this view, emphasising that AI is still a relatively new phenomenon and that people are only just learning to use it responsibly. Another senior practitioner added that they would not immediately fire someone for incompetent use of AI in a crisis, as the mistake might stem from factors other than a lack of fundamental communication skills (e.g., time pressure, too much trust placed in AI, or a lack of awareness of the risks). Thus, what matters is understanding *why* the mistake happened rather than rushing to make conclusions about the practitioner’s professional competence.

While mistakes may be forgiven in times of crisis, this forgiveness has clear limits. Practitioners drew a sharp distinction between crisis-driven errors and those that occur under normal working conditions. In a crisis, time pressure and chaos can explain why someone relies too heavily on AI without verifying its output. Outside a crisis, that same excuse, however, does not hold. As one practitioner put it, in routine situations, there is no time pressure, and people are expected to check AI outputs before using them. The expectation of professional diligence is higher precisely because the conditions allow for it. Another senior interviewee described having to fire an employee who repeatedly used AI carelessly in routine work and turned it into a persistent negative habit:

“I had in my own team one person with whom we had to have a more serious conversation on a similar topic [reckless AI use]. And in the end, I let them go. It was visible that their work came directly from chat [AI]. There was no depth. There were mistakes inside.”

This quote illustrates that when careless AI use becomes a pattern outside of crisis conditions, it is no longer treated as a forgivable mistake but as a failure of

professional judgement. In such cases, Duff's (2009, p. 978) conditions for fair blame are more likely to be met because the practitioner has time to control the results of an AI tool and an opportunity to reflect on whether using it is appropriate.

5.3.2. Answerability: Managers in Charge

When asked who should explain AI-related failures, practitioners tend to point to managers rather than the individuals who made the mistake. Mostly, they expected communication heads to handle the situation. This shows an interesting understanding of how responsibility works in an organisational context. It does not matter who pressed the buttons or who entered the data into AI, what matters is who holds the official oversight role. This delegation of responsibility means that even if a junior employee made the mistake, it is mostly their manager who must face the public and provide an explanation. A senior interviewee explained:

“The idea of organisation is precisely that there would exist control layers and distribution of responsibility. Usually, at least two pairs of eyes should see the text before it goes to the public. If a junior employee had the mandate to send a text out without review, then finally responsible is the one who gave them that mandate.”

This thought corresponds to Smith (2015, p. 103), who argues that answerability can exist independently of the person who *caused* the problem. The interviews show that seniors are held answerable based on their job role, not because they personally chose to use AI. This is further supported by Van de Poel's (2011, p. 38) distinction between responsibility-as-task and responsibility-as-authority. Responsibility-as-task concerns the responsibilities that come with your job, for example, a bus driver is responsible for driving the bus because that is what they do (Van de Poel, 2011, p. 38). Responsibility-as-authority means being in charge of something: if one is responsible for a project, one has authority over it and everything that happens within it (Van de Poel, 2011, p. 38). In this example, managers of the communication team have both. They are answerable because of their status (they have authority over their team) and because of their role (overseeing the organisation's communication is their job). Thus, answerability gets assigned by position, and even when a junior employee makes a mistake using AI, the manager must still answer for it because their authority over the team makes them answerable for everything that happens under their watch.

Nevertheless, when asked how leaders should explain AI-related mistakes to the public, practitioners offered markedly different answers, revealing uncertainty about whether AI involvement should be publicly disclosed. Some saw the use of AI as increasingly normal and not worth hiding. One pointed out an example of a government official who had publicly stated in a radio interview that AI drafts documents for him. The same interviewee argued that using AI has become an expected and normal practice not only in communication but in other fields as well. Thus, from this view, there is no need to feel embarrassed when discussing the use of AI, as it is just another modern tool in the workflow. Yet others took a more cautious position: if, in a crisis, journalists or stakeholders directly asked whether AI was used, they would not lie but would immediately shift the focus away from AI. There were also practitioners who argued for the avoidance of mentioning AI. A thought illustrated it:

“It would not go out in style “we analysed wrongly with AI” but rather “the information has been specified, here is the new and correct information”. [...] I would not emphasise at all that it [use of AI] was a mistake.”

Such different positions are not just personal preferences but also reveal professional uncertainty about whether AI use signals either 1) competence and being a modern organisation with a tech-driven mindset or 2) incompetence and lack of human care. This uncertainty reflects what Johnson and Verdicchio (2017, p. 2268) identify as “AI anxiety”, meaning there is unease among practitioners about whether using AI strengthens or undermines professional credibility. In other words, there is still no clear consensus on how transparent organisations should be regarding their use of AI. Some see this as a sign of progress and an opportunity to showcase the modern tools they use, while others fear it could undermine trust.

5.3.3. Accountability: The “Non-Negotiable” Human Oversight

When discussing blame, error, or consequences, which Watson (1996, p. 230) refers to as accountability, practitioners were in complete agreement on one point: AI itself can never be blamed for wrongful outcomes. This may seem self-evident, but it is important because it shows where practitioners draw a line: no matter how much AI contributed to a problematic outcome, no matter how opaque the “black box” is, and no matter how much pressure people were under when using it, the question of blame, in their opinion, always falls on humans. Two participants explained this as follows:

“Take self-driving cars as an example. If a car hits a person and the person is injured, [...] who is truly responsible? The computer does not make that person healthy again. [...] It [AI] is not a moral agent. Responsibility does not disappear.”

“With AI, the validation duty is on the human. Today, there is no AI that can be blindly trusted. It is like if a three-year-old child does stupid things, is the child guilty or the mother? A three-year-old is so dumb that they do not understand what they do. Parents are there to monitor and guide. AI is the same: it can do things and sometimes very well, but in the end, you do not leave a three-year-old alone at home.”

This aligns with Anderson and Anderson’s (2007, p. 19) view that moral responsibility requires intentionality and free will, which AI lacks. The child metaphor is an interesting way to illustrate how practitioners understand the validation duty with AI: it can “do things” and sometimes “very well”, yet cannot be blamed for wrong outcomes because it does not possess moral intelligence. When looking beyond the communication field, a similar understanding is further supported regarding self-driving vehicles, where responsibility cannot be attributed to the technology itself, but must instead rest with the human and institutional frameworks involved in design and use (Cunneen et al., 2019, p. 722).

Nevertheless, it can be argued that this “human oversight” way of thinking assumes that people can fully understand and oversee how AI tools work. In reality, this is often not the case. As discussed earlier in Chapter 5.1.3, many AI systems function as “black boxes”, meaning their decision-making processes are difficult to interpret (Buhmann, 2023, p. 3; Hermann, 2022, p. 1269). As a result, even when there is an expectation that humans control AI’s outputs, this is not always straightforward in practice. To deal with this issue, one expert suggested that accountability should rather depend on the informed choice of using AI:

“A person makes a conscious decision to put information there [into AI] to get something back [...]. That is their decision. And if a person makes a conscious decision, they must take responsibility for it.”

The thought above is a great example of Hedlund and Persson’s (2025, p. 4056) idea of contribution-based responsibility: those who decide to insert data into AI bear responsibility for the effects and own the consequences. Tigard (2021, p. 441), however, challenges this view, arguing that holding a single person accountable for AI’s actions is questionable. The interviews reflected a similar confusion around accountability.

Rather than placing blame on the individual user, one interviewee argued that accountability should instead lie with the organisation that permitted the use of AI, and that if something goes wrong, the whole organisation should be held accountable. Another practitioner, in contrast, narrowed the focus to managers or other similar senior roles who allow or promote the use of AI:

“If it is created [a “sandboxed” version of AI] for our internal use and someone has encouraged to use it and guaranteed that it is safe, then it is no longer only my responsibility. It is a bit like a doctor who says: “Take five tablets a day”. If that is wrong guidance and the person takes them, then the responsibility is on the doctor. [...] In such a situation the one who has directed to use or made the decision [...] is responsible.”

Although Tigard (2021, p. 441) sees no problem with distributing accountability among multiple actors, it still does not entirely resolve the question of who specifically should be held accountable if mistakes occur. When responsibility is spread across several parties, it becomes unclear who is to blame and what consequences should follow, which only reinforces the responsibility gap discussed in Chapters 5.1.4 and 5.2.3. For example, when sensitive information is leaked, it may be difficult to determine whether the blame lies with the practitioner who entered the data, the management team that approved or oversaw the system, or the technical teams that developed the AI tool. In a crisis, this kind of uncertainty and confusion can make an organisation seem less transparent and less in control. This is particularly concerning because stakeholder trust relies heavily on transparency and plays an important role in organisational legitimacy and flexibility (Zerfass & Viertmann, 2017, p. 75). Since AI is already often associated with concerns about trust and transparency (Zerfass et al., 2020, p. 385), unclear accountability can further damage the relationships organisations rely on during crises.

6. Concluding Discussion

This chapter presents a concluding discussion of the analysis. The research problem stems from a practical dilemma: AI has introduced new ethical and responsibility questions into crisis communication, where professionals must answer for messages shaped by systems they cannot fully understand. The aim is to provide new knowledge and enhance the understanding of how communication professionals in Estonia understand and negotiate questions of ethical AI use and responsibility when using AI tools in crisis communication. Specifically, how they describe AI's role, draw boundaries between human and AI tasks, and determine responsibility when AI-supported communication leads to problematic outcomes.

Firstly, the analysis reveals four themes regarding AI's role in crisis communication: efficiency versus loss of skills, AI as a sensemaking support, the “black box” problem, and the absence of AI guidelines. Practitioners position AI as a tool, as does much of the literature (e.g., Cheng et al., 2024, p. 177), but Guzman and Lewis (2020, pp. 71–72) argue this view is naïve, since AI also participates in meaning-making rather than neutrally processing information. This study's participants reflect this duality: while they describe AI as a tool, several recognised that it actively shapes which aspects of a situation receive attention and that they cannot always detect when AI has steered their interpretation in a misleading direction. If practitioners cannot say with certainty where their own judgement ends and AI's influence begins, it becomes difficult to maintain the assumption that the user controls AI and is therefore fully responsible for the outcome.

Estonian practitioners navigate this tension by drawing a clear line: AI handles routine work, humans handle strategic thinking. This mirrors Zerfass et al.'s (2018, p. 495) distinction between operational and strategic communication⁴. By reserving strategic work for themselves, practitioners ensure they can explain and stand behind their decisions, which is what moral responsibility requires (Fischer &

⁴ Operational is “getting a situation right every time” through known procedures, strategic is “getting this right” in a unique, high-stakes moment where there is no blueprint (Zerfass et al., 2018, p. 495).

Ravizza, 1998, p. 2). AI is excluded from strategic tasks not mainly because it might produce errors, but because its involvement would undermine practitioners' ability to fully account for important decisions they are expected to justify. Nevertheless, as Zerfass et al. (2018, p. 495) further note, the boundary between operational and strategic is volatile: what looks routine can turn into a strategic disaster without warning. Therefore, a less important crisis communication task considered safe to delegate to AI may carry strategic consequences if the tool makes an error and the practitioner fails to notice it. Since excessive or careless use of AI can gradually weaken humans' ability to critically assess AI outputs (Vallor, 2015, p. 108), practitioners may become less likely to notice such errors over time.

Secondly, the analysis shows that all three criteria practitioners use to decide on task allocation, such as crisis severity, verification ability, and data sensitivity, serve the same underlying function. They protect the conditions under which moral responsibility can be maintained. Duff (2009, p. 978) argues that holding someone responsible requires both control and knowledge, and practitioners exclude AI precisely where these conditions break down: when the crisis is too severe to risk losing control, when the output is too complex to verify with their own knowledge, or when the data is too sensitive to expose. Therefore, in crisis communication, the boundary between AI and human tasks is shaped less by capability than by responsibility: practitioners do not base delegation only on what AI can do, but on whether they, humans, can still stand behind the result.

Nonetheless, it is worth considering why practitioners apply this "control and knowledge" standard so strictly to AI but not to human colleagues. A senior who delegates a situation analysis to a junior, in fact, faces the same problem they describe with AI because they cannot see how the junior reached their conclusions either. One explanation for this can be that human delegation preserves the possibility of dialogue: a junior can be asked how they arrived at a conclusion. AI cannot participate in that kind of dialogue, because each prompt gives a slightly different answer. Without access to the reasoning behind an output, the conditions for answerability (Smith, 2015, p. 103) are harder to maintain. How practitioners draw boundaries around AI is therefore not only about managing risk but about protecting their sense of control over their own professional judgement. Simultaneously, these boundaries are not entirely in practitioners' hands. Private organisations with "sandboxed" AI systems can delegate a wider range of tasks to AI, whereas public-sector practitioners face regulations

that restrict AI use regardless of its potential value. What practitioners can responsibly delegate to AI thus depends not only on their judgement, but on what their organisation makes possible for them.

Thirdly, the discussion of responsibility reveals an interesting contradiction: practitioners universally insist that humans must remain in control, yet when asked who, specifically, should face the consequences, their answers differ. Regarding attributability, crisis functions as a contextual buffer. Practitioners resist judging a colleague's professionalism based on AI mistakes made under extreme pressure, viewing such errors as products of the situation rather than reflections of character (Fischer & Ravizza, 1998, p. 2). This line of reasoning is understandable, but it also creates a contradiction: crisis situations, where the consequences of AI errors are likely to be the most severe, are precisely the situations in which those errors are judged most lightly.

When it comes to answerability, practitioners consistently shift this responsibility to managers, rather than to those who actually used the AI. This follows a well-established organisational logic (Van de Poel, 2011, p. 38) and, from a stakeholder perspective, is arguably a strength. A manager carrying the explanation signals that the organisation takes the issue seriously at a leadership level. Nevertheless, this arrangement also means that the actual learning from AI-related mistakes may not reach the right level internally. If the manager consistently takes answerability, the individual practitioner who made the error may face less pressure to reflect on their own AI use, which over time could weaken the critical self-awareness that responsible AI use demands.

This brings the discussion to its most practical concern. Even if blame is softened, explanation duties are delegated, and humans must oversee AI outputs, the question of who ultimately bears the consequences for AI errors remains. Regarding accountability, practitioners are united in saying AI itself cannot be blamed, but they are divided on which "human" should be. Some point to the individual who used AI, others to the manager who approved the work or to the organisation that implemented the tool. This echoes Tigard's (2021, p. 443) call for distributed responsibility, but as the analysis showed, distributing responsibility without a clear structure makes it harder to locate at all, which, in turn, directly undermines stakeholder trust and transparency that crises demand (Zerfass & Viertmann, 2017, p. 75).

On a broader level, the analysis shows that AI has not simply entered crisis communication as another tool to be managed but complicates the very foundations on which the field relies. Crisis communication involves advising organisations under uncertainty, time pressure, and heightened scrutiny (Coombs & Holladay, 2022, p. 259), and it requires clear professional judgement, ethical awareness, and the ability to take responsibility for one's decisions (Falkheimer et al., 2016, p. 150). Yet, what emerges from this thesis is that these 16 Estonian practitioners working in communication are still working out how AI fits into that picture: where moral boundaries are drawn individually because shared standards do not yet exist and where the principle of human responsibility is widely accepted but not yet clearly structured.

The issues and somewhat contradicting thoughts raised by practitioners in this thesis also extend beyond crisis communication and Estonia. In many fields, AI is being integrated into professional decision-making processes faster than moral, legal, and organisational frameworks can adapt to it (Cunneen et al., 2019, p. 721; Shad, 2025, p. 891). As a result, practitioners often rely on norms that were developed before AI became part of their work (Shad, 2025, p. 892). This reflects a state of anomie: a form of normative uncertainty in which existing norms no longer provide sufficient guidance for determining the ethical and acceptable use of AI in professional practice (Marks, 1974, p. 333). The way in which the 16 strategic communication practitioners in this thesis address these issues is just one example of this situation. Although the stakes in crisis communication may seem lower than in fields where AI can cause physical harm (such as weapons or vehicles), the consequences can still be severe. For example, a false statement could cause significant financial damage to a publicly traded company, mislead people involved in a dangerous crisis, or undermine a government's stability. Who bears responsibility when an AI system contributes to such a statement remains unresolved, much like in the cases of autonomous cars or weapons. Thus, if a profession built around helping others navigate uncertainty has not yet resolved where responsibility lies for its use of AI, it is reasonable to expect similar tensions in any field where AI is becoming part of consequential work. In that sense, this thesis contributes not only to strategic communication but to the broader societal discussion about the challenges AI poses across professions.

There are also reasons to argue that this uncertainty regarding AI will not resolve on its own. The analysis already shows that practitioners describe organisations as being under pressure to adopt AI simply to stay competitive. Interviews

indicated that those who use AI can work faster and at lower cost, which leaves little room for others to slow down and address ethical questions before adopting AI themselves. Simultaneously, AI systems are becoming more autonomous (Russell & Norvig, 2023, p. 20), meaning the human oversight that practitioners in this study see as essential will become harder to ensure as the technology advances. Another question concerns the next generation in strategic communication. If the operational tasks that have traditionally served as starting points for junior professionals (e.g., desktop research) are increasingly handled by AI, it is worth asking what entry-level work will look like in the future and whether enough of it will remain for newcomers to build their careers on.

Taken together, this discussion highlights that AI has introduced unresolved ethical questions and contradictions, as well as new workflows in crisis communication, which practitioners are still actively adapting to. As AI integration increases, the traditional rules and boundaries that have historically guided professional judgement in strategic and crisis communication must now be renegotiated and reshaped (Bowen & Erzikova, 2025, p. 3). This leaves practitioners operating in an environment characterised by significant uncertainty, and to navigate this shift, the field of strategic communication must make targeted efforts to resolve these tensions and provide practitioners with clearer guidelines for using AI in crisis communication.

6.1. Contributions to Research and Practice

This thesis contributes to research and practice in several ways. Firstly, applying the three “faces” of moral responsibility (Shoemaker, 2011, p. 603; Watson, 1996, p. 229) to AI use in crisis communication shows that practitioners do not treat responsibility as a single unified concept. Instead, they reason differently depending on what is at stake: whether a mistake says something about someone’s competence, who should explain what went wrong, or who should bear the consequences. Existing research about AI ethics in communication often discusses responsibility without making these distinctions, thereby risking the failure to see how each dimension of moral responsibility operates according to its own logic. Separating them makes visible how each dimension works in practice (e.g., that crisis pressures managers to excuse employees’ mistakes rather than judge their character, or that the person who must explain a failure is determined by organisational position rather than by who actually used AI).

These patterns would remain hidden if moral responsibility were treated as a one-dimensional concept.

Furthermore, Guzman and Lewis (2020, pp. 71–72) have argued that AI participates in meaning-making rather than passively transmitting information. This study contributes empirical evidence for that claim by showing that in crisis communication, AI does not simply assist practitioners but actively shapes which aspects of a situation receive attention. This means that when organisations decide which crisis tasks to delegate to AI, they should also consider how it may influence practitioners' understanding of the crisis itself.

Regarding practice, the findings suggest that organisations should take a more active role in governing the use of AI in (crisis) communication rather than leaving individual practitioners to navigate it on their own. Developing guidelines that specify which tasks are appropriate for AI, how outputs should be verified, and how responsibility is distributed when errors occur would provide the structural clarity that is currently missing in many organisations. Additionally, AI literacy programmes that extend beyond technical proficiency to include awareness of AI's limitations and ethical questions, such as hallucination and opacity, would better prepare practitioners to use AI responsibly in high-stakes situations. Additionally, organisations should establish clear review processes for AI-assisted crisis communication, ensuring that outputs are assessed not only by communication professionals but also by subject-matter experts. Together, these steps would help ensure that ethical AI use and responsibility in crisis communication are not carried by practitioners alone but supported by the organisational structures within which they work.

6.2. Recommendations for Future Research

This thesis used a fictive but realistic crisis vignette to prompt reflections on AI use and responsibility. While this generated valuable insights into how practitioners reason through these questions, it remains hypothetical, as the emotions and consequences of a real crisis may lead to different reasoning. Future research could examine how responsibility is understood in actual crisis situations where AI was used. As AI becomes more integrated into communication work, cases of problematic AI-assisted crisis outcomes will likely become more common, offering researchers concrete material to investigate.

Additionally, AI tools and their integration into communication practice are evolving rapidly. Therefore, longitudinal research tracking how practitioners' perceptions of AI, their task delegation practices, and understanding of responsibility develop over time would offer insight into whether the moral deskilling concern raised by this thesis's participants, but also by Vallor (2015, p. 108), will become an even bigger concern as AI use becomes increasingly common. Finally, this study examined responsibility solely from the practitioners' perspective, yet the analysis repeatedly showed that stakeholder perceptions of AI use, especially in tragic crises, shape how practitioners draw ethical boundaries. While previous research has already looked into stakeholder reactions to AI-generated communication (e.g., Kirk & Givi, 2025, p. 6; Schneider & Boman, 2025, p. 146), these studies focus on perceptions of individual messages rather than responsibility. Future research could therefore investigate how stakeholders assign responsibility when they learn AI was involved in crisis communication and how this affects their trust in the organisation.

References

- Alvesson, M. (2023). *Interpreting Interviews* (2nd ed.). SAGE Publications.
- Alvesson, M., & Kärreman, D. (2011). *Qualitative Research and Theory Development: Mystery as Method*. SAGE Publications.
- Anderson, M., & Anderson, S. L. (2007). Machine Ethics: Creating an Ethical Intelligent Agent. *AI Magazine*, 28(4), 15–26.
<https://doi.org/10.1609/aimag.v28i4.2065>
- Berger, P., & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Doubleday.
- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative Research for Education: An Introduction to Theory and Methods* (2nd ed.). Boston: Allyn & Bacon.
- Bory, P., Natale, S., & Katzenbach, C. (2024). Strong and Weak AI Narratives: An Analytical Framework. *AI & Society*, 40, 2107–2117.
<https://doi.org/10.1007/s00146-024-02087-8>
- Bowen, S. A. (2024). “If it can be done, it will be done”: AI Ethical Standards and a Dual Role for Public Relations. *Public Relations Review*, 50(5), Article 102513. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Bowen, S. A., & Erzikova, E. V. (2025). Introduction: Defining futurism in strategic communication: needs and directions. In S. A. Bowen & E. V. Erzikova (Eds.), *Handbook of Innovations in Strategic Communication* (pp. 2–12). Edward Elgar Publishing. <https://doi-org.ludwig.lub.lu.se/10.4337/9781035326488>
- Brinkmann, S., & Kvale, S. (2015). *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd ed.). SAGE Publications.
- Buhmann, A. (2023). The Implications of Artificial Intelligence for Strategic and Organisational Communication. *Communication & Organisation*, 64(2), 1–7. <https://doi.org/10.4000/communicationorganisation/13066>
- Buhmann, A., Zerfass, A., Laborde, A., Moreno, A., Romenti, S., Tench, R., & Siegel, C. (2026). “We’re right in the foothills”: The adoption of

- artificial intelligence in corporate communication departments. *Corporate Communications: An International Journal*, 1–21.
<https://doi.org/10.1108/CCIJ-09-2025-0296>
- Burr, V. (2015). *Social Constructionism* (3rd ed.). Routledge.
- Cheng, Y., Lee, J., & Qiao, J. (2024). Crisis Communication in the Age of AI: Navigating Opportunities, Challenges, and Future Horizons. In W. T. Coombs (Ed.), *Media and Crisis Communication* (1st ed., pp. 172–192). Routledge. <https://doi.org/10.4324/9781032619187>
- Coombs, W. T. (1995). Choosing the Right Words: The Development of Guidelines for the Selection of the “Appropriate” Crisis-Response Strategies. *Management Communication Quarterly*, 8(4), 447–476.
<https://doi.org/10.1177/0893318995008004003>
- Coombs, W. T. (2006). The Protective Powers of Crisis Response Strategies: Managing Reputational Assets During a Crisis. *Journal of Promotion Management*, 12(3/4), 241–260. https://doi.org/10.1300/J057v12n03_13
- 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., & Holladay, S. J. (2022). Crisis communication as strategic communication: Process and insights. In J. Falkheimer & M. Heide (Eds.), *Research Handbook on Strategic Communication* (pp. 259–273). Edward Elgar Publishing. <https://doi.org/10.4337/9781800379893.00025>
- Creswell, J.W. (2007). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Cunneen, M., Mullins, M., & Murphy, F. (2019). Autonomous Vehicles and Embedded Artificial Intelligence: The Challenges of Framing Machine Driving Decisions. *Applied Artificial Intelligence*, 33(8), 706–731.
<https://doi.org/10.1080/08839514.2019.1600301>
- Damar, M., Özen, A., Çakmak, Ü. E., Özoğuz, E., & Erenay, F. S. (2024). Super AI, Generative AI, Narrow AI and Chatbots: An Assessment of Artificial Intelligence Technologies for The Public Sector and Public Administration. *Journal of AI*, 8(1), 83–106. <https://doi.org/10.61969/jai.1512906>

- Duff, R. A. (2009). Legal and Moral Responsibility. *Philosophy Compass*, 4(6), 978–986. <https://doi.org/10.1111/j.1747-9991.2009.00257.x>
- Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). From the Editors: Grand Challenges and Inductive Methods: Rigor without Rigor Mortis. *Academy of Management Journal*, 59(4), 1113–1123. <https://doi.org/10.5465/amj.2016.4004>
- Ertel, W. (2017). *Introduction to Artificial Intelligence* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-319-58487-4>
- European Commission. (2019, April 8). *Ethics Guidelines for Trustworthy AI*. Publications Office of the European Union. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Falkheimer, J. (2021, September 29). Legitimacy Strategies and Crisis Communication. In E. Hannah (Ed.), *Oxford Research Encyclopedia of Politics*. Oxford University Press. <https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-1567>
- Falkheimer, J., & Heide, M. (2006). Multicultural Crisis Communication: Towards a Social Constructionist Perspective. *Journal of Contingencies and Crisis Management*, 14(4), 180–189. <https://doi.org/10.1111/j.1468-5973.2006.00494.x>
- Falkheimer, J., & Heide, M. (2022). *Strategic Communication: An Introduction to Theory and Global Practice* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003168997>
- Falkheimer, J., & Heide, M. (2023). Strategic Improvisation in Crisis Communication. In W. T. Coombs & S. J. Holladay (Eds.), *The Handbook of Crisis Communication* (pp. 405–420). John Wiley & Sons. <https://doi.org/10.1002/9781119678953.ch27>
- Falkheimer, J., Heide, M., Simonsson, C., Zerfass, A., & Verhoeven, P. (2016). Doing the Right Things or Doing Things Right? Paradoxes and Swedish Communication Professionals' Roles and Challenges. *Corporate Communications: An International Journal*, 21(2), 142–159. <https://doi.org/10.1108/CCIJ-06-2015-0037>
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

- Flick, U. (2023). *An Introduction to Qualitative Research* (7th ed.). SAGE Publications.
- Frandsen, F., & Johansen, W. (2011). The Study of Internal Crisis Communication: Towards an Integrative Framework. *Corporate Communications: An International Journal*, 16(4), 347–361.
<https://doi.org/10.1108/13563281111186977>
- Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829–839. <https://doi.org/10.2307/2023833>
- Franklin, S. (2014). History, Motivations, and Core Themes. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 15–33). Cambridge University Press.
- Giddens, A. (1999). Risk and Responsibility. *Modern Law Review*, 62(1), 1–10.
<https://doi.org/10.1111/1468-2230.00188>
- Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*, 5(1), 1–46.
<https://doi.org/10.2478/jagi-2014-0001>
- Government Office of Estonia. (n.d.). *Eesti.ai initiative*. Retrieved May 17, 2026, from <https://www.riigikantselei.ee/en/supporting-government-and-prime-minister/eestiai-initiative>
- Gregory, A., & Smith, B. G. (2025). AI State of Play: The Technology, Strategic Communication Practice and Academia. *International Journal of Strategic Communication*, 19(2), 119–129.
<https://doi.org/10.1080/1553118X.2025.2474875>
- Gunkel, D. J. (2012). Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century. *Communication +1*, 1(1), 1–26.
<https://doi.org/10.7275/R5QJ7FR>
- Guzman, A. L., & Lewis, S. C. (2020). Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda. *New Media & Society*, 22(1), 70–86. <https://doi.org/10.1177/1461444819858691>
- Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>

- Hall, J. S. (2011). Ethics for Machines. In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 28–44). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511978036.005>
- Hallahan, K., Holtzhausen, D., van Ruler, B., Verčič, D., & Sriramesh, K. (2007). Defining Strategic Communication. *International Journal of Strategic Communication*, 1(1), 3–35.
<https://doi.org/10.1080/15531180701285244>
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication*, 25(1), 89–100.
<https://doi.org/10.1093/jcmc/zmz022>
- Hao, X., Demir, E., & Eyers, D. (2025). Beyond Human-in-the-Loop: Sensemaking Between Artificial Intelligence and Human Intelligence Collaboration. *Sustainable Futures*, 10, Article 101152.
<https://doi.org/10.1016/j.sftr.2025.101152>
- Hearit, K. M. (2006). *Crisis Management by Apology: Corporate Response to Allegations of Wrongdoing*. Routledge.
<https://doi.org/10.4324/9781410615596>
- Hedlund, M., & Persson, E. (2025). Distribution of Responsibility for AI Development: Expert Views. *AI & Society*, 40, 4051–4063.
<https://doi.org/10.1007/s00146-024-02167-9>
- Hermann, E. (2022). Artificial Intelligence and Mass Personalization of Communication Content: An Ethical and Literacy Perspective. *New Media & Society*, 24(5), 1258–1277.
<https://doi.org/10.1177/14614448211022702>
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Science and Engineering Ethics*, 21(3), 619–630. <https://doi.org/10.1007/s11948-014-9565-5>
- Hossain, M. Z., Akter, N., Hasan, L., Bepari, M., & Sultana, S. (2025). Artificial Intelligence and Machine Learning in Crisis Communication: A Management Information System Perspective. *European Journal of Innovative Studies and Sustainability*, 1(3), 149–163.
[https://doi.org/10.59324/ejiss.2025.1\(3\).12](https://doi.org/10.59324/ejiss.2025.1(3).12)

- Jaques, T. (2007). Issue Management and Crisis Management: An Integrated, Non-linear, Relational Construct. *Public Relations Review*, 33(2), 147–157. <https://doi.org/10.1016/j.pubrev.2007.02.001>
- Jenkins, N., Bloor, M., Fischer, J., Berney, L., & Neale, J. (2010). Putting It in Context: The Use of Vignettes in Qualitative Interviewing. *Qualitative Research*, 10(2), 175–198. <https://doi.org/10.1177/1468794109356737>
- Johnson, D. G. (2015). Technology with no Human Responsibility? *Journal of Business Ethics*, 127(3), 707–715. <https://doi.org/10.1007/s10551-014-2180-1>
- Johnson, D. G., & Verdicchio, M. (2017). AI Anxiety. *Journal of the Association for Information Science and Technology*, 68(9), 2267–2270. <https://doi.org/10.1002/asi.23867>
- Kaplan, A. (2021). Artificial Intelligence (AI): When Humans and Machines Might Have to Coexist. In P. Verdegem (Ed.), *AI for Everyone? Critical Perspectives* (pp. 21–32). University of Westminster Press. <https://doi.org/10.16997/book55.b>
- Kirk, C. P., & Givi, J. (2025). The AI-authorship Effect: Understanding Authenticity, Moral Disgust, and Consumer Responses to AI-generated Marketing Communications. *Journal of Business Research*, 186, Article 114984. <https://doi.org/10.1016/j.jbusres.2024.114984>
- Levine, S. (2023, February 22). Vanderbilt apologizes for using ChatGPT in email on Michigan shooting. *The Guardian*. <https://www.theguardian.com/us-news/2023/feb/22/vanderbilt-chatgpt-ai-michigan-shooting-email>
- Loof, T. (2025). Listening and AI. In R. Luttrell & A. A. Wallace (Eds.), *Public Relations and the Rise of AI* (pp. 166–183). Routledge. <https://doi.org/10.4324/9781032671482>
- Marks, S. R. (1974). Durkheim's Theory of Anomie. *American Journal of Sociology*, 80(2), 329–363. <http://www.jstor.org/stable/2777505>
- Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology*, 6(3), 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- McCarthy, J., & Hayes, P. J. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. L. Webber & N. J. Nilsson

- (Eds.), *Readings in Artificial Intelligence* (pp. 431–450). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-934613-03-3.50033-7>
- Moor, J. H. (1985). What is Computer Ethics? *Metaphilosophy*, 16(4), 266–275. <https://doi.org/10.1111/j.1467-9973.1985.tb00173.x>
- Moor, J. H. (2011). The Nature, Importance, and Difficulty of Machine Ethics. In M. Anderson & S. L. Anderson (Eds.), *Machine Ethics* (pp. 13–20). Cambridge University Press. <https://doi.org/10.1017/CBO9780511978036.003>
- Morse, S. J. (2004). Moral and Legal Responsibility and the New Neuroscience. In J. Illes (Ed.), *Neuroethics: Defining the Issues in Theory, Practice, and Policy* (2nd ed., pp. 33–50). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198567219.003.0003>
- Muggleton, S. (2014). Alan Turing and the Development of Artificial Intelligence. *AI Communications*, 27(1), 3–10. <https://doi.org/10.3233/AIC-130579>
- Nies, H., & Zhao, L. (2025). AI-Powered Enhancements in Media Relations: Exploring Potential and Ethical Considerations in Public Relations. In R. Luttrell & A. A. Wallace (Eds.), *Public Relations and the Rise of AI* (pp. 118–135). Routledge. <https://doi.org/10.4324/9781032671482-10>
- Panda, G., Upadhyay, A. K., & Khandelwal, K. (2019). Artificial Intelligence: A Strategic Disruption in Public Relations. *Journal of Creative Communications*, 14(3), 196–213. <https://doi.org/10.1177/0973258619866585>
- Patton, M. (1990). *Qualitative Evaluation and Research Methods*. Beverly Hills, CA: Sage.
- Piller, E. (2025). Inhuman Rhetoric: Generative AI and Crisis Communication. *Journal of Business and Technical Communication*, 39(1), 42–50. <https://doi.org/10.1177/10506519241280594>
- Prahl, A., & Goh, W. W. P. (2021). “Rogue Machines” and Crisis Communication: When AI Fails, How do Companies Publicly Respond? *Public Relations Review*, 47(4), Article 102077. <https://doi.org/10.1016/j.pubrev.2021.102077>
- Prasad, P. (2018). *Crafting Qualitative Research: Beyond Positivist Tradition* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315715070>
- Rashid, A. B., & Kausik, M. A. K. (2024). AI Revolutionizing Industries Worldwide: A Comprehensive Overview of its Diverse Applications. *Hybrid*

- Advances*, 7, Article 100277. <https://doi.org/10.1016/j.hybadv.2024.100277>
- Rich, E. (1985). Artificial Intelligence and the Humanities. *Computers and the Humanities*, 19(2), 117–122. <https://www.jstor.org/stable/30204398>
- Russell, S. J., & Norvig, P. (2023). Artificial Intelligence: Definition and Background. In H. Sheikh, E. Schrijvers, & C. Prins (Eds.), *Mission AI: The New Frontier of Artificial Intelligence in Government* (pp. 15–39). Springer. <https://doi.org/10.1007/978-3-031-21448-6>
- Santoni de Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them. *Philosophy & Technology*, 34, 1057–1084. <https://doi.org/10.1007/s13347-021-00450-x>
- Sarker, I. H. (2022). AI-based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, 3, Article 158. <https://doi.org/10.1007/s42979-022-01043-x>
- Schneider, E. J., & Boman, C. D. (2025). Digital Dilemmas: AI's Influence on Ethical Crisis Communication. In R. Luttrell & A. A. Wallace (Eds.), *Public Relations and the Rise of AI* (pp. 136–148). Routledge. <https://doi.org/10.4324/9781032671482>
- Schoenherr, J. R., & Thomson, R. (2024). When AI Fails, Who Do We Blame? Attributing Responsibility in Human–AI Interactions. *IEEE Transactions on Technology and Society*, 5(1), 61–70. <https://doi.org/10.1109/TTS.2024.3370095>
- Seidenglanz, R., & Baier, M. (2023). The Impact of Artificial Intelligence on the Professional Field of Public Relations/Communications Management: Ethical Issues, Challenges, and an Attempt at a Forecast. In A. Adi (Ed.), *Artificial Intelligence in Public Relations and Communications: Cases, Reflections, and Predictions* (pp. 124–136). Quadriga University of Applied Sciences.
- Sellnow, T. L., & Seeger, M. W. (2013). *Theorizing Crisis Communication*. Wiley-Blackwell.

- Shad, K. B. (2025). Artificial Intelligence-Related Anomies and Predictive Policing: Normative (Dis)orders in Liberal Democracies. *AI & Society*, 40, 891–902. <https://doi.org/10.1007/s00146-023-01751-9>
- Shoemaker, D. (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics*, 121(3), 602–632. <https://doi.org/10.1086/659003>
- Smith, A. M. (2015). Responsibility as Answerability. *Inquiry*, 58(2), 99–126. <https://doi.org/10.1080/0020174X.2015.986851>
- Steyn, B. (2003). From Strategy to Corporate Communication Strategy: A Conceptualisation. *Journal of Communication Management*, 8(2), 168–183. <https://doi.org/10.1108/13632540310807402>
- Steyn, B. (2018, August 22). Communication Strategy. In R. L. Heath & W. Johansen (Eds.), *The International Encyclopedia of Strategic Communication*. Wiley-Blackwell. <https://doi.org/10.1002/9781119010722.iesc0036>
- Sturges, D. L. (1994). Communicating through Crisis: A Strategy for Organizational Survival. *Management Communication Quarterly*, 7(3), 297–316. <https://doi.org/10.1177/0893318994007003004>
- Sullins, J. P. (2011). When is a Robot a Moral Agent? In M. Anderson & S. L. Anderson (Eds.), *Machine Ethics* (pp. 151–161). Cambridge University Press. <https://doi.org/10.1017/CBO9780511978036.013>
- Sutherland, K. E. (2025). *Artificial Intelligence for Strategic Communication*. Palgrave Macmillan. <https://doi.org/10.1007/978-981-96-2575-8>
- Swiatek, L., & Galloway, C. (2022). Artificial Intelligence and Public Relations: Growing Opportunities, Questions, and Concerns. In D. Pompper, K. R. Place, & C. K. Weaver (Eds.), *The Routledge Companion to Public Relations* (pp. 352–362). Routledge. <https://doi.org/10.4324/9781003131700>
- Talbert, M. (2024, June 3). Moral Responsibility. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2025/entries/moral-responsibility/>

- Tigard, D. W. (2021). Artificial Moral Responsibility: How We Can and Cannot Hold Machines Responsible. *Cambridge Quarterly of Healthcare Ethics*, 30(3), 435–447. <https://doi.org/10.1017/S0963180120000985>
- Tracy, S. J. (2020). *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact* (2nd ed.). John Wiley & Sons.
- Trapp, R. (2025, April 26). What Tiny Estonia Can Teach Others About Digital Success. *Forbes*. <https://www.forbes.com/sites/rogertrapp/2025/04/26/what-tiny-estonia-can-teach-others-about-digital-success/>
- Vahtla, A. (2017, September 5). Potential security risk could affect 750,000 Estonian ID cards. *ERR News*. <https://news.err.ee/616732/potential-security-risk-could-affect-750-000-estonian-id-cards>
- Vallor, S. (2015). Moral Deskillling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy & Technology*, 28(1), 107–124. <https://doi.org/10.1007/s13347-014-0156-9>
- Van de Poel, I. (2011). The Relation Between Forward-Looking and Backward-Looking Responsibility. In N. A. Vincent, I. van de Poel, & J. van den Hoven (Eds.), *Moral Responsibility: Beyond Free Will and Determinism* (pp. 37–52). Springer.
- Volk, S. C., & Buhmann, A. (2023). Digital Corporate Communication and Measurement and Evaluation. In V. Luoma-aho & M. Badham (Eds.), *Handbook on Digital Corporate Communication* (pp. 118–133). Edward Elgar Publishing. <https://doi.org/10.4337/9781802201963.00>
- Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. <https://doi.org/10.2478/jagi-2019-0002>
- Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, 24(2), 227–248. <https://www.jstor.org/stable/43154245>
- Weick, K. E. (1988). Enacted Sensemaking in Crisis Situations. *Journal of Management Studies*, 25(4), 305–317. <https://doi.org/10.1111/j.1467-6486.1988.tb00039.x>
- Zerfass, A., & Viertmann, C. (2017). Creating Business Value through Corporate Communication: A Theory-Based Framework and its Practical Application. *Journal of Communication Management*, 21(1), 68–81. <https://doi.org/10.1108/JCOM-07-2016-0059>

- Zerfass, A., Hagelstein, J., & Tench, R. (2020). Artificial Intelligence in Communication Management: A Cross-National Study on Adoption and Knowledge, Impact, Challenges, and Risks. *Journal of Communication Management*, 24(4), 377–389. <https://doi.org/10.1108/JCOM-10-2019-0137>
- Zerfass, A., Verčič, D., Nothhaft, H., & Werder, K. P. (2018). Strategic Communication: Defining the Field and Its Contribution to Research and Practice. *International Journal of Strategic Communication*, 12(4), 487–505. <https://doi.org/10.1080/1553118X.2018.1493485>
- Zerfass, A., Buhmann, A., Laborde, A., Moreno, A., Romenti, S., & Tench, R. (2024, November). *European Communication Monitor 2024. Managing Tensions in Corporate Communications in the Context of Geopolitical Crises, Artificial Intelligence, and Managerial Learning*. European Public Relations Education and Research Association (EUPRERA). <https://www.akademische-gesellschaft.com/wp-content/uploads/2024/11/ecm-2024-25-report.pdf>
- Zhao, H., Falkheimer, J., & Heide, M. (2017). Revisiting a Social Constructionist Approach to Crisis Communication – Investigating Contemporary Crises in China. *International Journal of Strategic Communication*, 11(5), 364–378. <https://doi.org/10.1080/1553118X.2017.1363758>

Appendix 1. Interview Guide

Introduction

Thank you for agreeing to participate in this interview and in my master's thesis on the questions of responsibility and ethics of artificial intelligence in crisis communication. The interview will last 45–60 minutes and will be audio recorded for transcription and analysis. All information will be treated confidentially, and your identity and organisation will be protected. Before this interview, you received an informed consent form. Do you have any questions before we begin? If everything is clear, I will now start the recording.

Warm-up and professional background (5-8 minutes)

- Could you briefly describe your current role and main responsibilities?
- What is your professional background?
- How long have you been working in communication-related roles?
- Have you been involved in crisis or high-pressure communication situations?
- If yes: could you briefly describe the nature of those situations?
- Does your organisation have any guidelines or policies regarding the use of AI in (crisis) communication? Please describe it to me.
- If not, how have you thought through the use of AI in your every-day work?

Crisis vignette: Your organisation is responsible for a widely used digital identification platform in Estonia that is relied on daily for banking, healthcare access and signing official documents. Late in the afternoon, you are informed about a possible security incident. At this stage, it is unclear what exactly happened, but there is a risk of a cyberattack originating from a foreign eastern country. Hackers may have accessed personal data of up to 400,000 residents, including pin-codes for the identification platform, personal ID numbers, document numbers, and ID card photos.

Technical teams are still investigating and cannot yet confirm the scope of the breach. Simultaneously, journalists begin requesting comments, and social media

discussions escalate quickly, with speculation about identity thefts. Screenshots and unverified claims start circulating online. Senior management asks you and your communication team to closely monitor public reactions, support leadership decision-making, and prepare a first public statement within the next hour. The message is expected to calm public concern while avoiding statements that could later turn out to be incorrect.

RQ1: role and use of artificial intelligence in crisis communication (15 minutes)

- Please walk me through how you would handle this situation step by step, from the moment you become aware of the incident until the first statement is published.
- At which points in this process, if any, would you consider using AI tools?
- If you used AI in this situation, what and how would you ask it to do (for example, what prompt would you type)?
- What would you mainly use AI for at those points, and what would you rely on your own judgement for?
- How does the use of AI in this kind of crisis differ from its use in routine communication work, in your experience?
- In this process, where do you think AI adds value, and where does it not? Please elaborate your answers.

RQ2: appropriateness and task allocation (20 minutes)

- Looking back at the scenario but also from your professional experience, which specific tasks in crisis communication would you consider appropriate to assign to AI?
 - Why specifically these tasks?
- Which tasks should clearly remain human-led?
 - Why?
- What considerations guide your decision when deciding whether to involve AI or not?
- What information would you be willing to insert into an AI system during a crisis?

- What information would you avoid inputting?
 - Why?
- Does the severity or sensitivity of the crisis affect where you draw these boundaries between human and AI involvement? If yes, how do you draw the boundaries? What primarily guides this decision?
- Can you please describe to me, in your opinion, what makes a crisis communication task “too risky” to delegate to AI?

RQ3: responsibility in AI-supported crisis communication (15 minutes)

Developments regarding the vignette: In the first hour after the incident becomes public, your team prepares an initial public message under strong time pressure. The technical team shares a short internal update with the communication team, but the information is highly technical and still uncertain. To communicate quickly and clearly, a communication practitioner from your team uses ChatGPT to summarise the technical update into a short public-facing statement and then edits it briefly before publication. The summary appears professional and plausible, and because of severe time pressure the practitioner does not double-check every detail against the technical report before publication. The public statement is as follows: “We are currently investigating a potential security incident affecting the digital identification system and are working closely with relevant authorities. Based on the information available at this stage, the incident appears limited and there is currently no evidence of misuse of citizens’ data. We will share further updates as soon as more verified information becomes available.”

The message is published on the organisation’s official channels and states that the incident appears limited and that there is currently no evidence of data misuse. However, over the following days it becomes clear that this wording was misleading, because personal data had in fact been accessed and affected citizens should have taken precautionary steps early on.

In hindsight, the organisation is criticised for publishing information that was not fully verified and for communicating too much reassurance too early. Journalists and stakeholders also question whether the organisation relied too heavily on AI

tools and whether the organisation acted transparently once the mistake became clear.

- Looking at the published message and evolved situation, what is your professional reaction to it?
- Attributability: what, if anything, does this situation reveal about the professional judgement of the people involved in drafting and approving the message?
- Answerability: who should be expected to explain or justify why the message was communicated in this way, given the uncertainty and time pressure?
 - Why them?
- Accountability: do you think it would be fair to criticise or blame anyone for this communication outcome?
 - Why or why not?
- What kind of consequences would be appropriate (if any)?
- Does it matter for responsibility that the message was produced through an AI summary of a technical report and was not checked line-by-line before publication?
 - If yes: how does it change your view?
- Once it becomes clear that the first message may have been misleading, what would be the responsible way to correct it publicly?
- Would you mention that AI was used?
 - Why or why not?

Conclusion (5 minutes)

- Is there anything we have not discussed that you consider important for understanding AI, responsibility and crisis communication?
- Is there anything you would like to clarify or add?
- May I contact you again if follow-up questions arise?

Thank you very much for your time and participation.

Appendix 2. Consent Form

Participation in a Master's Thesis Study on Estonian Communication Practitioners' Perspectives on Ethical AI Use and Responsibility in Crisis Communication

This master's thesis explores how communication professionals in Estonia use artificial intelligence (AI) tools in crisis communication and how they understand ethical AI use and negotiate responsibility when AI is part of the communication process. The study focuses on professional decision-making in situations where communication must be produced quickly and where messages may be ethically sensitive due to public scrutiny, uncertainty, or the emotional impact of the crisis.

The thesis seeks to map how practitioners describe the role of AI in crisis communication, how they decide which crisis communication tasks are appropriate to assign to AI, and how responsibility is understood when AI-supported communication contributes to unintended, contested, or problematic outcomes. The study does not evaluate individual participants or organisations. Instead, it aims to develop academic knowledge about professional reasoning and responsibility in AI-supported communication work.

The study is based on individual, semi-structured interviews with communication professionals who have practical experience using AI tools in their work. The interview will last approximately 45–60 minutes and will be conducted online (Teams/Google Meet). The interview will focus on professional experiences, practices, and reflections. Participants are not expected to disclose confidential or sensitive organisational information.

With the participant's consent, the interview will be audio recorded and later transcribed for research purposes. The interview transcript will be analysed as part of the thesis. The collected empirical material will be used in the analysis section of the master's thesis to illustrate research analysis. Access to raw empirical material (audio

recordings and transcripts) will be only for researcher and the supervisor. The empirical material will be stored securely and will be deleted after the thesis has been completed and approved.

Confidentiality is ensured. Participants' names, workplaces, and other identifying details will not be included in the thesis. If necessary, details will be generalised or removed to prevent identification. Any third parties mentioned during the interview will be removed. The completed thesis may be made publicly available through the university's electronic repository.

Participation in the study is voluntary. Participants may decline to answer any question and may withdraw from the study at any time without providing a reason and without any negative consequences.

Consent:

I, _____, confirm that I have been informed about the purpose and content of the study and understand what participation involves, including the interview format and approximate time commitment. I understand that participation is voluntary and hereby give my consent to take part in the interview.

I understand that I may withdraw from the study at any time and may refuse to answer any question. I also understand that I may request additional information or clarification from the researcher if needed.

Researcher's name: Karmen Kikas

Participant's name: XX

Researcher's signature: XX

Participant's signature: XX

Date: XX

Date: XX