

Crisis Communication in the Age of Deepfakes: A Qualitative Analysis of Government and Media Response to AI-Manipulated Media

Emma Ransom Smith

University of Memphis

Abstract: This study examines crisis communication in an era of deepfakes through a qualitative analysis of the 2022 deepfake video showing Ukrainian President Volodymyr Zelenskyy announcing the surrender of Ukrainian forces. The study investigates how Zelenskyy's communication team responded to the fabricated video, how international media organizations covered and verified the incident, and how government and media responses contributed to the reconstruction of trust, authenticity, and credibility. Drawing on Situational Crisis Communication Theory (SCCT) and Media Credibility Theory (MCT), the study uses a qualitative case study design combining documentary analysis and thematic analysis. The government uses official statements, presidential video addresses, press releases, and verified social media communications issued between March 2022 and October 2023. The media analysis examines purposively selected coverage from Reuters, BBC, VBC, Al Jazeera, and The Washington Post, focusing on verification framing practices, source attribution, credibility, and technological and geopolitical framing. The findings show that the Ukrainian government primarily focused on rapid denial, corrective communication, and an authentic counter message, which was delivered through verified channels. International media organizations complemented the efforts through visible verification practices, including comparison with authentic footage and identification of visual inconsistencies, source evaluation, and corroboration. The study argues that credibility in deepfake crises is therefore coproduced among governments, media organizations and technology platforms. The use of SCCT and MCT theories shows a more comprehensive framework for understanding these crises while highlighting the need to move from traditional repair toward the management of authenticity, institutional trust and evidentiary transparency.

Keywords: Deepfakes, Authenticity, Ukraine, Misinformation, Crisis.

INTRODUCTION

The study's central argument is that the effectiveness of crisis communication in the age of deepfakes depends largely on an institution's ability not only to communicate an appropriate crisis response but also to establish the authenticity of its communication channel and provide credible verification of its messages. Using the framework of Situational Crisis Communication Theory (SCCT) and Media Credibility Theory, the study seeks to establish how governmental communicators and international news organisations reconstruct trust, credibility, and authenticity when AI-manipulated media compromises the evidentiary value of political audiovisual communication. Methodologically, the study will adopt a qualitative case-study design, employing documentary analysis and qualitative thematic analysis of official Ukrainian government communications and selected international news coverage of the Zelenskyy deepfake event. Adopting this approach, the study positions deepfakes not merely as a new category of misinformation but as a distinctive crisis-communication phenomenon that challenges the correlation between media verification, source authority, and public trust

The Zelenskyy deepfake demonstrates this problem with clarity. The concocted video sought

to communicate an extraordinary political decision directly from Ukraine's President during wartime. Its relevance thus hinged on the credibility normally attributed to Zelenskyy's verbal and visual presence. Government communicators had to rapidly counter the false message, while journalists had to authenticate the video and frame or flag it for audiences. The case ultimately raises suspicions that existing crisis communication research does not sufficiently answer: how should crisis-response strategies be adapted when a communication's source has been impersonated, what credibility cues become decisive when visual evidence is unreliable, and how news organisations and governments collectively reestablish confidence in authentic information?

Moreover, relatively few studies have explicitly integrated SCCT with media credibility outlooks to analyse how credibility and trust are reestablished after an AI-mediated communication attack. This theoretical separation limits understanding of deepfake crises since the phenomenon simultaneously consists of response strategy, crisis responsibility, audience trust and confidence, source credibility, and journalistic verification. The current study hence aims to address this gap by analysing the Zelenskyy deepfake as a case of technologically facilitated

crisis communication where the authenticity of the communication channel itself was contested. This study examines how President Zelensky's communication team responded to the deepfake, how international media outlets also covered and framed the incident, and how the government of Ukraine and media applied SCCT and Media Credibility Theory to restore trust, authenticity and credibility.

Theoretically, the study contributes considerably to the scholarship on crisis communication by extending SCCT into a communication setting where the crisis may arise from the fabrication of an organisation's communicative identity instead of from traditional organisational misconduct or operational failure. By integrating SCCT with MCT, this study also links two bodies of scholarship that are mostly analysed separately. This integration will give a nuanced explanation of how response strategy, responsibility, verification, and response strategy interact during AI-arbitrated crises. Empirically, the study contributes critically to the growing scholarship on political deepfakes by evaluating an actual crisis rather than depending solely on hypothetical or experimental scenarios. The Zelenskyy case is especially valuable since it happened during an active international conflict and consisted of governmental communication, social-media platforms, journalism, and cyber-enabled distribution. Analysing the responses of these actors can thus reveal insightful dimensions into the practical dynamics of managing deepfake crises.

As news organizations increasingly rely on digital audiovisual material, the ability to validate content has become a primary aspect of journalistic credibility. Recent verification practices increasingly consist of source evaluation, cross-platform comparison, metadata examination, and technological analysis. For instance, Reuters has described dedicated visual-verification processes that are communicated to authenticate user-generated material (Reuters, 2026). Understanding how such verification processes are communicated to audiences is significant since credibility depends not solely on identifying false content but also on proving why audiences must trust the verification process.

Practically, the study may be useful to governmental communication professionals, journalists, public relations practitioners, and media organisations in devising more resilient

crisis communication mechanisms. Deepfake situations require speed, but speed without verification can worsen misinformation. On the other hand, delayed correction may allow manipulated content to gain traction and perceived authenticity. The study therefore reveals evidence about the communicative practices through which institutions can respond quickly while showing authenticity simultaneously.

LITERATURE REVIEW

Crisis communication emerged from the pragmatic need for organisations and public institutions to communicate under conditions of uncertainty, intense public scrutiny, and threat. Although contemporary crisis communication is now a recognised academic discipline integrating theories from public relations, communication, management, psychology, and political communication, its historical emergence is closely linked with the professionalization of media relations and public information. In a landmark study, Frandsen & Johansen (2020) assert that crisis communication developed through the interrelation between practical organisational responses to crises and the later institutionalisation of crisis management as an academic field.

One crucial historical account is Ivy Lee's response to the Pennsylvania Railroad disaster in 1906. The intervention by Lee is often associated with the shift towards more direct and open communication between journalists and organisations. Instead of depending solely on secrecy or withholding information after a crisis, Lee's approach spotlighted the provision of factual information to journalists and the public. The relevance of this historical event lies less in the conjecture that Lee invented crisis communication in its entirety than in the fact that it demonstrates an early paradigm shift towards proactive, organised, and information-driven communication during crises (Frandsen & Johansen, 2020).

The historical development of crisis communication can ultimately be interpreted as a paradigm shift from control of information, through management of information, towards competition over information authenticity. This distinction is key to understanding deepfake crises. In a conventional crisis, an organisation might try to correct or control information about an incident. In a deepfake crisis, however, the organisation may need to show that a seemingly authoritative communication is not authentic. The challenge is

then simultaneously technological, communicative, and epistemological.

This evolution also explains why recent crisis communication cannot be explored solely as reputation management. The discipline has increasingly acknowledged that crisis communication performs relational, informational, and behavioural functions. During a crisis, audiences need information that helps them to understand what has happened, decide what actions they should take, and evaluate whether the institution responsible for responding to the crisis is trustworthy. This is especially significant in governmental crises since the effects of inaccurate information can transcend reputational damage to political stability, public safety, and institutional legitimacy.

The theoretical development of crisis communication has generated many impactful frameworks for explaining and understanding how organisations respond when their legitimacy or reputation is endangered. Of the lot is Benoit's Image Repair Theory (IRT). Although IRT originated before the period covered by this literature review, its continuing importance is shown by more recent empirical and theoretical scholarship. In his foundational works, Benoit (2020, 2021) describes image repair as "a process through which individuals or organisations respond to criticism, accusations, or perceived wrongdoing that threatens their image". IRT identifies five comprehensive strategies: denial, evasion of responsibility, minimising offensiveness, corrective action, and mortification.

The relevance of IRT for modern crisis communication lies in the sense that it conceptualizes crisis response as primarily rhetorical. The organisation does not just transmit information; it formulates a defence orchestrated to influence how relevant audiences interpret the responsibility and character of the organisation. This makes the theory particularly relevant for analysing political communication, where governments and leaders often defend their legitimacy and authenticity under circumstances of public scrutiny.

Contemporary studies affirm the continuing usefulness of IRT. For instance, studies evaluating Malaysian government communication during the COVID-19 pandemic revealed evidence of all five main image repair strategies in governmental media statements, with minimising offensiveness

and corrective action strategies among the most frequently adopted strategies. The study also indicated a correlation between image repair strategies and public opinion, proving that IRT is still applicable to governmental crisis communication instead of being limited to corporate reputation management (Masngut & Mohamad, 2021). In the same vein, Ferguson *et al.* (2018) revealed that public relations practitioners distinguished among image repair strategies according to crisis circumstances, implying that practitioners do not essentially perceive all strategies as equally suitable.

Irrespective of its continuous relevance to contemporary crisis communication management and scholarship, IRT has shortcomings when applied to a deepfake crisis. The theory's main positioning is towards the rhetorical defence of an actor implicated in wrongdoing. A deepfake causes a considerably different circumstance since the said act may not have happened at all. In the case of Zelenskyy's deepfake crisis, for example, the main communicative challenge was not just that Zelenskyy had been implicated in making an undesirable statement. The primary problem was that a manipulated audiovisual representation was being attributed to him. The communicative task ultimately comprised authenticating the source, disproving the message, as well as rebuilding confidence in the legitimacy of the communication channel.

SCCT offers compatible development, improving understanding and broadening usefulness. Coombs' SCCT theorizes crisis response according to the connection between crisis type, perceived responsibility, and reputational risk. SCCT draws considerably on attributional reasoning and suggests that the perception of stakeholders about responsibility significantly shapes the extent of reputational risk experienced by an organisation and thus the suitability of diverse response strategies (Coombs, 2020). SCCT distinguishes between accidental, victim, and preventable crisis clusters, and relates these to different levels of attributed responsibility.

This distinction is especially significant for a deepfake implicating a top government actor or official. A government targeted by a concocted video may be understood as a victim of a malevolent act and not considered the perpetrator of the apparent communication. Eventually, denial, informational responses, and corrective action become particularly important. The Ukrainian

government did not need to apologise for Zelenskyy's purported surrender since no such incident (surrender) actually happened. Rather, the government was supposed to affirm that the apparent communication was false and fraudulent and give audiences reliable alternative evidence.

The relevance of SCCT in recent times is noticeable in empirical research applying the framework to digitally fabricated crises. For instance, Kwok *et al.* (2022) applied SCCT to investigate crisis messages on social media and revealed how organisations adapt message types to crises. Again, studies consisting of hundreds of thousands of organisational tweets during the COVID-19 pandemic indicated that adjusting and instructing information could minimise reputational threats in victim crises and persuade stakeholder reactions (Dhar & Bose, 2022). These studies validate the hypothesis that SCCT can operate effectively in digitally fabricated settings.

SCCT has also been applied to explicitly political crises. Studies comparing the crisis communication strategies of Donald Trump and Andrew Cuomo during the COVID-19 pandemic identified that political actors did not automatically conform to the response strategies stipulated by SCCT. Trump largely adopted diminishment and denial, while Cuomo stressed greatly on cohesion and bolstering (Wang *et al.*, 2022). This finding is theoretically significant since it shows that political communication may complicate the supposition that communicators simply choose the response strategy most closely related to crisis responsibility.

Situational Crisis Communication Theory: Assumptions, Contributions, and Criticism

SCCT rests upon multiple significant assumptions that require critical examination in the context of deepfakes. The first is that stakeholders are capable of making at least a provisional attribution of crisis responsibility. The theory assumes that audiences evaluate whether an organisation is accountable or responsible for a crisis and that these perceptions impact reputational threat. This is reasonable in conventional crises comprising accidents, misconduct, product failures, or natural disasters. Nonetheless, a deepfake may intentionally manipulate the attribution processes themselves.

The second assumption is about the existence of an identifiable crisis incident. SCCT requires a researcher to classify the crisis and then determine

the appropriateness of the response. In a deepfake crisis, a researcher must initially differentiate between the 'represented event' and the 'actual event.' The deepfake creates a seemingly real event that is ontologically not true. Ultimately, the crisis is partially generated by the distribution of the representation itself.

The third assumption is that communication can be treated as a response to a crisis instead of considering it as part of the crisis environment. In a deepfake event, however, communication is simultaneously the object of fabrication and the main mechanism through which the manipulation is opposed. This leads to a recursive problem: the government must use mediated communication to convince audiences that mediated communication has been fabricated.

SCCT is also criticised for its historical emphasis on reputation and organisations. Governments differ from commercial organisations since their legitimacy or authenticity cannot be limited to consumer attitudes or corporate reputation. Also, governmental crisis communication consists of political authority, national identity, public safety, and institutional legitimacy. Political crisis communication can thus contain competing audiences with diverse expectations and political interpretations.

A further limitation regards the role of technological affordances. SCCT was not developed particularly to account for AI-generated content, platform moderation, algorithmic amplification, or digital forensic verification. Recent studies increasingly point out that AI is becoming an active aspect of crisis communication. In a systematic review of 177 SSCI-indexed studies, Cheng *et al.* (2025) identified that AI-related crisis communication literature has expanded considerably but remains scattered. Significantly, the review documents that research has disproportionately emphasised classification and detection rather than trust-building and relational functions, while relatively little research explicitly adopts theoretical frameworks.

Deepfake as a Crisis Communication Phenomenon

Deepfakes are a representation of a particular category of synthetic media created or fabricated using machine learning technologies. Their importance lies not merely in their ability to create false information but in their capacity to simulate

the identity and communicative behaviour of actual individuals. Early literature revealed deepfakes as a potential threat to privacy, democracy, and national security (Chesney & Citron, 2019). The concern was that fabricated media could be employed to concoct political statements, create confusion about real events, and control public opinion. Comparably, Westerlund (2019) identified journalism, political communication, and warfare as areas particularly susceptible to deepfake technologies. Also, more recent literature has moved beyond hypothetical warnings towards empirical assessment. For instance, Vaccari & Chadwick (2020) found that exposure to political deepfakes can amplify uncertainty and erode trust in news, even when audiences do not essentially accept the manipulated content as true. This finding is key to the present study since it transitions the analytical focus from deception to uncertainty. A deepfake does not need to convince everyone to be strategically effective. If it forces audiences to question whether a government statement is genuine, whether a news report has been fabricated, or whether visual evidence can still be trusted, it may attain a destabilising impact without creating widespread belief in the false content.

Contemporary scholarship further complicates the belief that deepfakes are inherently more persuasive than other forms of misinformation. In that regard, Hameleers *et al.* (2024) found that political deepfakes containing plausible fabrications can be regarded as more credible, whereas implausible deepfakes may nonetheless possess delegitimizing impacts on political actors. This difference is crucial since it points out that political harm and credibility are not identical outcomes. Groh *et al.* (2024) give additional empirical evidence that humans have difficulty in differentiating between certain political speech deepfakes, especially when imitated audio is created by advanced text-to-speech technologies. The experimental evidence of the authors proves that audiovisual cues can help with detection or identification, but the wider implication is that technological advancements may progressively undermine the reliability of unaided human verification.

The Zelenskyy Deepfake and the Russian Ukrainian Information

The March 2022 deepfake about Ukrainian President Zelenskyy offers a particularly important case study for analysing these theoretical concerns. During the Russian invasion of Ukraine, a

fabricated video showed the President instructing Ukrainian forces to surrender. The video circulated after the Ukrainian media infrastructure was compromised and subsequently disseminated via online platforms. The importance of the incident goes beyond the technical quality of the fabricated video. The incident happened during an active armed conflict between Russia and Ukraine, a period when information was closely connected to troop morale, strategic communication, and political authority. The manipulated message thus sought to appropriate the authority of an incumbent head of state at a period when his words carried considerable psychological and political relevance. Academic studies on the incident ascertain that the case offers one of the earliest real-world examples of deepfake deployment in warfare. Twomey *et al.* (2023) argue that the Russian-Ukrainian war offers a particularly valuable context for studying the epistemic consequences of deepfakes since the conflict created several examples of manipulated media about political leaders.

The case also shows the significance of pre-bunking and institutional preparedness. The Ukrainian authorities had warned audiences about the possibility of fabricated communications before the deepfake appeared. This is theoretically significant because it confuses the traditional temporal distinction between pre-crisis and crisis communication. In traditional crisis management, preparation takes place before the event, while response follows the crisis. In a deepfake setting, nonetheless, pre-crisis communication can create the credibility infrastructure needed for later authentication. The Zelenskyy case thus allows for analysis of the correlation between prior credibility and crisis response. A government that has created recognised communication channels, reliable patterns, and consistent visual identities of direct communication may have greater capacity to prove authenticity during a fabricated media attack.

Media Credibility Theory (MCT) and Trust in Digital Environments

Media credibility concerns the perceived accuracy, reliability, trustworthiness, and fairness of the sources of information and media channels. Early credibility studies affirmed the significance of source characteristics in determining the effectiveness of communication, while later studies expanded the frontier to technological and media environments. In modern digital communication, credibility cannot be explored and understood only as a property of the communicator

since audiences increasingly assess the interaction among message, source, context, and medium. Metzger & Flanagin (2013) conceive online credibility evaluation as consisting of cognitive heuristics through which users assess information in complex digital settings. Although their work predates the current generation of generative AI, its conceptual relevance has increased tremendously as fabricated media has made source and content more challenging. The major relevance of MCT to deepfakes lies in their theory's acknowledgement that audiences use several cues when deciphering whether information should be trusted. These may consist of source reputation, corroboration, consistency, platform characteristics, presentation, and external validation. A deepfake attacks many of these cues at the same time. For instance, a manipulated video may mirror the appearance of a legitimate political leader, thus generating a false source cue. It may also appear on a legitimate platform, generating a false medium cue. Again, a deepfake may contain a probable political statement, giving a contextual cue. The resulting consequence is thus not based on a single false claim but on the fabrication of the credibility infrastructure surrounding the message.

This explains why deepfakes challenge the conventional heuristic of "seeing is believing." Similarly, Shin & Lee (2022) give empirical evidence that deepfake content can shape perceptions of news credibility, while Groh *et al.* (2024) reveal the continuing difficulty humans experience when differentiating between certain manipulated political speech from authentic speech. The problem becomes even more daunting when false content is placed within credible news environments. For example, a journalist may ignorantly report fabricated footage, or a legitimate media organisation may publish a correction after originally reproducing a false narrative. In such situations, the credibility of the original content and the credibility of the organisation distributing the deepfakes become interrelated. The comprehensive study conducted by Tandoc Jr *et al.* (2021) offers evidence for this institutional dimension. Their study found that misinformation can have impacts that extend beyond individual content to the information institution itself. This finding is validated by studies on deepfake-specific credibility. For instance, Ahmed *et al.* (2023) found that exposure to deepfakes can decrease positive perceptions of news accuracy, and Shin & Lee (2022) identified that the influence

of deepfake exposure interacts with already existing attitudes. Considered collectively, these findings point out that credibility is not a stable attribute possessed by a media institution. It is contextual, relational, and potentially fragile.

Deepfakes and Institutional Trust

Institutional trust is particularly significant in modern scholarship given that government and media organisations occupy diverse but interrelated positions within the crisis-information ecosystem. Governments communicate authoritative information, whereas news institutions perform contextualization, verification, and dissemination functions. Contemporary studies increasingly exhibit that deepfakes can influence trust in these organisations. To confirm this, a recent study by Neyazi *et al.* (2024), analysing perceptions of deepfakes across Indonesia, Malaysia, Singapore, and India, revealed that trust in government influenced attitudes towards deepfakes. The findings of the authors are relevant because they show that deepfake perceptions are not decided solely by technological awareness; they are also integrated within already existing organisational relationships. This supports a relational understanding of deepfake credibility. Individuals who already trust government may interpret an apparent government correction differently from citizens who do not trust the government. On the other hand, citizens who trust highly in particular media organisations may be more likely to accept the organisation's verification or correction of a deepfake.

The scholarship therefore suggests that deepfake incidents operate within an already existing trust architecture. A crisis does not begin with a blank slate. Media institutions and governments get into crisis with reputational histories, existing political associations, and established credibility. The effectiveness of a response is likely to rely partly on these established conditions. This is a critical qualification to SCCT. SCCT is basically situational, but deepfake crises indicate that already existing institutional credibility may shape how the situation itself is understood and interpreted. The same denial message may be assessed diversely based on the perceived credibility of the institution issuing it. The challenge is particularly severe in politically tense environments. To demonstrate this, the study by Hameleers *et al.* (2024) explains that the effect of deepfakes can depend on the plausibility and political congruence of the fabricated content. In the same vein, Shin & Lee (2022) in their study

prove that already existing attitudes can impact vulnerability to deepfake news. These findings point out that political predispositions mediate credibility, rather than being established exclusively by objective evidence.

The empirical evidence above shows the implication for crisis communication is that governments cannot presume that a factual verification or correction will automatically reestablish trust. The correction itself must be disseminated through channels and practices that audiences regard as credible. This explains the importance of cross-verification by independent media, visible evidence of authenticity, and direct communication by political leaders.

Media Verification as a Credibility-Reconstruction Mechanism

One of the most important developments in the contemporary deepfake scholarship is the acknowledgement that journalism can function as a form of rebuilding credibility. Conventional journalism serves as an information gatekeeper, but in a manipulated media setting it increasingly performs an authentication function. The Zelenskyy case demonstrates this development. International media institutions reported that the video existed, confirmed it, investigated the authenticity of the video, compared it with original footage, contextualized the event, and consulted experts across the wider information environment. Such practices transform authentication or verification from an internal journalistic process into a visible communicative performance.

This visibility is theoretically relevant. If audiences cannot independently decide whether a video is fake or original, journalists may need to prove how they know that it is false. Credibility thus becomes partly procedural. The audience may trust a particular news institution because it identifies sources, explains the verification process, and shows consistency with independent evidence. However, the study by Twomey *et al.* (2023) proves that verification is not risk-free. Since audiences may not understand what constitutes a deepfake, excessive warnings about fabricated media can generate suspicion towards original footage. Media institutions experience a paradox: informing audiences about the threats of deepfakes can enhance media literacy while simultaneously promoting indiscriminate skepticism.

Assumptions and Limitations of Media Credibility Theory

First, MCT assumes that credibility is a perceptual judgement (Rahman, 2014). Meaning information can be objectively accurate but regarded or understood as untrustworthy, or objectively false but considered as credible. This differentiation is important to deepfake studies because the strategic impact of synthetic media relies on audience perceptions instead of objective falsity alone.

Second, media credibility literature can sometimes treat “the media” as a relatively coherent organisational category (Rahman, 2014). Modern information environments are more fragmented. Audiences encounter social media, government websites, influencers, mainstream news channels, algorithmically recommended content, and anonymous accounts within the same information ecosystem. In the end, credibility may relate to particular outlets, platforms, journalists, or personalities instead of “the media” as a whole.

Third, credibility scholarship assumes that audiences possess stable criteria (Metzger & Flanagin, 2013). Digital environments oppose this assumption. Individuals use diverse heuristics based on technological literacy, political identity, emotional involvement, and source familiarity.

METHODOLOGY

Research Design

The first-dimension analyses official communications issued by the Ukrainian presidency and government in response to the deepfake. The corpus involves official statements, presidential video addresses, press releases, and verified social media communications published from the emergence of the deepfake in March 2022 through October 2023. The documents were chosen because they directly addressed the deepfake, its implications, its authenticity, or the government’s response to the misinformation. The government corpus was analysed through qualitative textual analysis and thematic coding. Textual analysis was used to analyse tone, credibility claims, language, and rhetorical strategies integrated in the communications. Specific attention was given to the ways in which the government repudiated the originality of the fabricated video, assigned responsibility, gave corrective information, urged audiences to depend only on verified communication channels, as well as bolstered political authority and national resilience.

The coding framework was underpinned deductively by SCCT. Initial analytical categories included corrective action, denial, bolstering, attribution of responsibility, credibility reaffirmation, reassurance, and rebuilding (Coombs, 2007). Simultaneously, the analysis retained an inductive aspect so that themes not anticipated by SCCT could emerge directly from the data. Inductive codes like this included emotional tone, resilience, audience reassurance, national identity, and authentication cues. This combination of inductive and deductive coding is significant because the purpose of the study is not just to find out whether existing SCCT categories happened, but to investigate whether deepfake crises generate communication practices that transcend beyond conventional crisis response classification.

A formal coding framework was developed before the major analysis. The codebook specified the name of each code, its conceptual definition, inclusion and exclusion criteria, and examples of textual material that qualified for the code. The first deductive codes were derived from SCCT and MCT, while additional inductive codes were created during repeated engagement with the data. Where a new principle emerged, it was compared with existing codes before being integrated into the codebook. This process decreased conceptual overlap, ensuring that codes were applied consistently across the documentary corpus.

NVivo qualitative data analysis software was used to organise, code, and retrieve the textual data. The reference in the earlier methodology to “InVivo” refers to the in-vivo coding approach, instead of a separate software programme. NVivo was used as the software environment within which the documents were imported, organised, and coded. In-vivo coding was selectively used to preserve analytically significant terms, words, or phrases in the communicators’ or participants’ own languages, specifically where particular expressions expressed authenticity, credibility, reassurance, and resistance. NVivo thus supported coding, data management, comparison, and node organisation, but the interpretation of the data remained the obligation of the researcher. This means that the software was not used as an automated mechanism for identifying themes.

Since the study has a single primary coder, inter-coder reliability was not calculated. The study explicitly acknowledges this as a methodological limitation rather than as evidence of reliability that

was not established. However, qualitative validity was strengthened through the use of a formal codebook, repeated reading of the corpus, systematic application of coding rules, maintenance of an analytical audit trail, comparison of coded extracts across documents, and reflexive review of coding decisions. These protocols are consistent with the rationale that qualitative rigour should be highlighted through coherence, transparency, and reflexive engagement with the analytical process instead of mechanical application of quantitative reliability statistics alone (Braun & Clarke, 2021).

The second part of the methodology analyses how international media organisations covered, authenticated, and frame the president Zelenskyy’s deepfake. The study focuses on Reuters, BBC, VBC, the Washington Post, and AL Jazeera, consistent with the media corpus established in the study. Articles were deemed eligible for inclusion where they directly discussed the deepfake incident of President Zelenskyy, its circulation, technological implications, governmental response, verification, or broader geopolitical implication or relevance. The selection of media articles was done following a criterion-based purposive sampling approach, but not availability-based selection. Purposive sampling was adopted to enable the researcher to identify information-rich cases that directly help to answer the research questions (Patton, 2015). For each of the five media organisations, the researcher aimed to get between 6-10 articles, generating an anticipated corpus of about 30-50 articles.

The 6-10 range was not intended to consist of a statistically representative sample. Instead, it established a manageable qualitative corpus while allowing sufficient variation in verification practices, reporting, and framing. Articles were chosen using four criteria: first, articles published between March 2022 and October 2023; second, articles with direct reference to President Zelenskyy’s deepfake incident; third, articles that contain substantive analysis or reporting rather than merely reproduce the existence of the video; and fourth, articles that provide evidence relevant to at least one of the study’s analytical dimensions, namely source attribution, verification, credibility, crisis response, technological uncertainty, or geopolitical framing. Information sufficiency was used to determine the final number of articles within the 6-10 range, rather than by a claim of statistical saturation. Where additional articles considerably repeated already established themes

without contributing analytically different material, further selection from that media channel was discontinued. This consists of a pragmatic qualitative stopping rule based on thematic sufficiency and informational redundancy. It is thus more accurate to describe the sampling as a criterion-based, purposive strategy with an information-sufficiency stopping rule than as either a fixed quota or convenience sampling.

Media Coding and Analysis

The media corpus was analysed using thematic analysis, supported by NVivo. First, deductive categories were informed by MCT and consisted of trust, authenticity, and uncertainty. Additional inductive codes captured recurrent media frames coming from the articles, including technological threat, information warfare, hybrid warfare, propaganda, and strategic deception. The analysis was done through repeated reading of the selected articles, initial coding, comparison and consolidation of codes, development and refinement of themes, as well as the final thematic structure. Using Braun & Clarke's (2006) approach to thematic analysis offers the major procedural foundation, although their later work stresses the significance of treating themes as interpretive patterns instead of simply counting recurring phrases or words (Braun & Clarke, 2021).

A comparative analysis was later conducted across the five media channels. This comparison assessed similarities and distinctions in how journalists established the authenticity of falsity of the video, which institutions or experts they depended on, how they attributed responsibility, as well as how they situated the incident within broader technological and geopolitical contexts or narratives. The aim was not to rank the media channels according to credibility but to understand the communicative practices through which media credibility was created and communicated to audiences.

Data Analysis and Integration

The findings from the media and government components were integrated during the final stage of thematic analysis. Integration was necessary since governmental credibility and media credibility are mutually linked in a deepfake crisis: though government communications must build the authenticity of their own messages, journalists must also independently assess and communicate the evidentiary foundation for those claims.

SCCT was thus used mainly to interpret governmental crisis response strategies, including corrective action, denial, bolstering, and attribution, while MCT offered the main lens for examining source evaluation, evidence, verification, trust, and authenticity within media coverage. The two theories were later brought together to identify areas of divergence and convergence between governmental crisis communication and journalistic verification. The integrated analysis therefore examines three connected processes: how the government tried to rebuild confidence in its own communication; how media organisations developed the credibility of their reporting; and how these institutional responses collaborated to restore authenticity after the deepfake and disrupted confidence in audiovisual evidence.

Ensuring Qualitative Rigour

Due to the single-coder design, methodological transparency is particularly significant. The study thus does not claim inter-coder reliability. Rather, dependability and credibility are supported through a transparent sampling protocol, consistent inclusion and exclusion data, a formal codebook, repeated engagement with the data, documented coding decisions, and an audit trail of thematic development. Reflexivity is also integrated throughout the analysis to ensure that interpretations are based on the documentary evidence rather than imposed upon it. These strategies strengthen the confirmability and dependability of the analysis while recognising the inherent interpretive role of the researcher. The methodological objective is therefore not to remove interpretation, which would not be consistent with qualitative inquiry, but to make the interpretive process adequately systematic and transparent for readers to evaluate how the conclusions were deduced.

ANALYSIS AND DISCUSSION

Government Response: Denial, Correction, and Reassertion of Authenticity

The most notable government response, according to the articles sampled, was the rapid rejection of the fabricated message. This constituted the clearest application of the SCCT denial stance because the Ukrainian government was not responsible for the apparent incident. Rather, the President's identity had been appropriated to manufacture false communication. The government then positioned itself as the victim of the crisis, and not the source of the crisis. The

corrective response was particularly meaningful since it did not depend solely on a written denial. President Zelenskyy issued a legitimate video message through his verified and authentic channels. NPR reported that the President responded by stating categorically that: “We are defending our land, our children, our families. So, we don’t plan to lay down any arms until our victory” (Allyn, 2022). This was strategically crucial since the Ukrainian government countered a concocted audiovisual communication with another audiovisual communication whose source was more rapidly established.

The importance of this theme should be considered dominant in the government material because the main communicative task was to differentiate the authentic President from the synthetic representation. The response thus went beyond a traditional denial strategy. It was an act of authentication. President Zelenskyy’s genuine appearance established communication channels, location, and consistency with his previous wartime messages provided credibility cues that the manipulated video could not sufficiently reproduce.

This finding expands SCCT. Traditional denial strategies involve rejecting responsibility for an incident. In a deepfake crisis, regardless, denial also consists of rejecting the credibility or originality of the supposed communication itself. President Zelenskyy’s response consequently demonstrates a revised form of corrective action: the government did not say that the message was false; it highlighted authenticity through a verifiable and competing communication.

Attribution and the Information Warfare Frame

A second highly recurrent theme was about the positioning of the deepfake video within a broader information warfare context. The incident happened during Russia’s invasion of Ukraine, and Ukrainian authorities had already warned audiences regarding the possibility of fabricated audiovisual material. Recent reporting validates that Ukrainian authorities had already warned their citizens about the possibility of fabricated videos before the deepfake occurred. According to a report by Euronews, Ukraine’s Land Forces had advised citizens about how difficult it can be to differentiate such videos (deepfakes) from authentic videos or materials and cautioned that their purpose could be to “disorient, sow panic, disbelieve citizens and incite our troops to retreat.

Rest assured – Ukraine will not capitulate” (Holroyd & Olorunselu, 2022). The report further records President Zelenskyy saying, “the only ones who should give up arms are Russian soldiers” (Holroyd & Olorunselu, 2022). This pre-debunking and warning were critical since they altered the communicative meaning of the deepfake video that was later released. The event was not presented as a separate technological prank; rather, it was integrated into a broader narrative of information warfare and strategic deception.

The Ukrainian government framing was validated by international reporting. For instance, NPR reported that Ukrainian authorities had been cautioning their citizens for weeks about the possibility of Russia disseminating fabricated videos as part of its information warfare. This denotes an important point of convergence between government and media framing. Government officials interpreted the video as part of a belligerent information environment. Journalists, on the other hand, placed the incident within the wider information conflict surrounding the Russian invasion.

Media Verification as a Mechanism of Rebuilding Credibility

The most significant media-side theme was verification, which can be categorised as highly recurrent/dominant since the reporting did not only repeat the existence of the video. Journalists evaluated the authenticity of the video via visual inconsistencies, source attribution, comparison with genuine footage, and corroborating evidence. A Reuters report reproduced by the *Indian Express* gives a particularly clear example. The report noted that internet users immediately identified “discrepancies between the skin tone on Zelenskyy’s neck and face, the odd accent in the video, and the pixelation around his head (Reuters, 2022). The report consequently changed verification from an invisible newsroom activity into evidence communicated to the audience. This is very important per the proposition of the MCT. Credibility was not determined just by the authority of Reuters or another media outlet. Credibility was built by explaining why the content was considered inauthentic. The audience was given observable cues through which the conclusion could be examined. Also, The *Washington Post* similarly highlighted the poor technical quality of the video, pointing out that President Zelenskyy appeared “mostly motionless” and that his head appeared disproportionately large

compared to his body size (The Washington Post, 2022). Media reports like these highlight what can be described as evidentiary journalism: the media outlets did not just tag the deepfake (video) as false but offered reasons supporting the classification.

Resilience, Bolstering, and the Reassertion of Political Authority

A further highly recurrent theme was the use of resilience and political continuity to counter the psychological meaning of the deepfake. The *Washington Post* offers particularly relevant evidence of this continuity. Before and around the deepfake incident, the newspaper reported that President Zelenskyy had continually communicated directly with Ukrainians via Telegram. In an earlier message, President Zelenskyy vehemently rejected surrender narratives and categorically stated: “We won’t lay down our arms. We will defend our state” (The Washington Post, 2022). This communication history was important because the audience did not encounter the deepfake in isolation. It opposed an already trusted communicative pattern. The *Washington Post* later suggested that the Ukrainian government’s response proved that deepfakes could be “disarmed” when handled rationally and calmly. The relevance of this observation is that credibility was partly collective. The effectiveness of the corrective communication relied on the existence of a trusted and recognisable communication identity before the attack.

Media Framing of Technological and Geopolitical Threat

The media material (deepfake video) contained a recurrent technological threat frame. Reuters noted that the video was poorly edited yet warned that experts considered it a possible precursor to more complex forms of manipulation and deception. The same concern appears in academic studies. Vaccari & Chadwick (2020) posit that political deepfakes can cause uncertainty and erode trust in news even where audiences do not necessarily believe or trust the manipulated content. Similarly, Twomey *et al.* (2023) highlighted that the Ukrainian conflict offers a critical background for assessing the epistemic consequences of deepfakes. The media coverage thus goes beyond the immediate incident into a broader discussion about the trustworthiness or reliability of audiovisual evidence. This theme also generates a distinction between media and government framing. Government communication was necessarily aimed at immediate correction, reassurance, and resistance. By contrast,

journalistic coverage could take advantage of the deepfake event to discuss the long-term implications of manipulated media, including the probability that future deepfakes might be substantially more sophisticated and convincing. The government sought to close the credibility crisis, and journalism also aimed at explaining why the crisis is relevant beyond the immediate incident.

Government and Media Findings in Direct Dialogue: Co-Produced Credibility

The media and government thus validated each other most clearly around three concerns, which are the falsity of the deepfake, the significance of rapid verification, as well as the wider information-warfare context. The intersection is highlighted by NPR’s reporting that Zelenskyy’s original and real response, as well as the earlier warnings of the Ukrainian government, helped minimise the effect of the manipulated video. However, there was a crucial divergence concerning attribution. Government and Ukrainian sources had stronger incentives to associate the deepfake incident with hostile information warfare, with the accounts of responsible journalists distinguishing between evidence proving that the video was concocted and evidence showing exactly who created and disseminated the video. This distinction strengthens rather than erodes media credibility since it highlights epistemic restraint.

The findings hence highlight that credibility during a deepfake crisis is co-produced and not simply transferred. Government communication offers authoritative information and an original reference point; journalism offers independent contextualization and verification. Neither actor completely substitutes for the other. Government authority with no external verification may be seen as self-interested, whereas media verification with no access to original official communication may not have an authoritative counter-source.

DISCUSSION

Implications for Situational Crisis Communication Theory (SCCT) and Media Credibility Theory (MCT)

The findings offer empirical support for integrating SCCT and MCT. On one hand, SCCT explains why the Ukrainian government adopted denial, corrective actions, and bolstering strategies. Further, the government was considerably a victim of manipulated communication and not the perpetrator of the deepfake crisis. On the other

hand, MCT highlights why denial alone would have been inadequate to explain the crisis. The corrective message needed credible sourcing, observable evidence, and corroboration.

This is in line with Metzger & Flanagin's (2013) argument that audiences use more than one cognitive heuristic to evaluate information in an online environment. It also resonates with Flanagin & Metzger's (2007) observations on verification behaviours and message characteristics when evaluating credibility in an online environment. In the Zelenskyy case, credibility was therefore built up with a series of cues: known means of communication, consistency of all messages received, other visual inconsistencies in the fake, independent journalism, and the action of platforms. The results also support Vaccari & Chadwick's (2020) premise that uncertainty hurts its own. The deepfake did not convince most viewers that Zelenskyy had capitulated to make the connection. It was also a test in another sense, as to whether audiences could tell the truth from the lie in presidential communication. The evidence lends support to the study's overall theoretical claim that reputation management has to shift from the traditional approach to managing a reputation to managing authenticity. The communicative task isn't just about the reputation of an institution following an event. It's meant as a way to show that the institution actually said what it says that it said—or, in this case, didn't say what the 'fabricated' media seems to claim. The results show that SCCT and MCT are complementary, but insufficient explanations of credibility management in deepfake crises. While SCCT explains how governments can play the role of a victim, it is less able to explain how the authenticity of other media is established independently in the fragmented digital environment. MCT covers issues of source credibility, verification, corroboration, and trust, but offers less analysis about how governments strategically build crisis responses, ascribe malicious intent, and build audiences on the eve of an attack. Credibility in deepfake crises does not only stem from the institutional response, but also from journalistic verification, as the integration indicates.

The Zelensky case also provides key conditions to be considered in the process of generalisation. Ukraine's response was aided by the high visibility of a political leader, the presence of international media relations, and significant international media attention. However, the speed with which

Zelenskyy's real counter-message was spreading and the level of journalistic scrutiny it received are not necessarily easily replicable by less well-resourced governments or actors in less well-informed environments. But the case illustrates a process that can be generalized, namely, speedy denial along with independently verifiable evidence, and the effectiveness of that process is context-dependent.

The findings additionally have implications for platform governance. Removing the manipulated video from platforms like Facebook, YouTube and Twitter was part of a larger credibility reconstruction process—it was not a one-off technological fix. Shared responsibility between governments, journalists and platforms is therefore crucial for effective deepfake crisis management, which entails having timely authenticating information, independent verification and contextualization of competing information by media organisations and providing proportionate detection, labelling and removal systems by platforms. This entails that deepfake governance is a problem of communication across institutional and technological actors, through which credibility is co-produced.

CONCLUSION

The goal of this study was to explore the concept of crisis communication in the era of deepfakes by analyzing the responses of governmental and international media to the deepfake of Zelenskyy's surrender from 2022. The results show that the incident was not just a typical case of misinformation, but also a crisis of communicative authenticity, in which the identity of a legitimate political actor was usurped to produce what seemed like legitimate information. The dominant response strategy was the Ukrainian government's quick denial, as well as corrective communication and the implementation of an authentic counter-message. It was further strengthened by reinforcement and restatements of political continuity, which challenged not only this false claim, but the psychological narrative of Ukrainian surrender.

The media response was mainly based on evidence verification. International news organisations did not just call the video a hoax; they shared glaring inconsistencies, contrasted the video with real footage, and provided corroborating evidence. This changed the nature of verification from a journalistic practice inside to one that is visible; it became a way of building credibility. The

evidence together shows that credibility was co-produced; that is, government communication gave a credible and authoritative authenticity, while independent journalism gave further verification and contextualization. The study concludes that conceptual changes are needed in conventional crisis communication theory when the organisation is a victim of impersonation via AI. SCCT is useful because it provides an explanation of how the Ukrainian government has positioned itself as the victim, how it denies responsibility, what corrective action it has taken, and how it attributes and bolsters itself. But the central epistemic issue which cannot be accounted for by SCCT is why audiences should accept the government's denial given that the medium in which it is delivered is easily manipulated.

However, Media Credibility Theory fails to sufficiently explain the process by which an institution chooses and adopts an appropriate crisis-response posture because of the influence of source characteristics, corroboration, verification, and contextual cues on the credibility of corrective information. The integration of these, as such, offers a more robust explanatory model. The main theoretical conclusion of the study is that the attitude of reputation repair is changing to managing authenticity and reconstructing trust. During deepfake crises, institutions need to regain not only their image but the trust in the authenticity of their communication network.

Governments should build pre-crisis authentication infrastructures made up of authenticated communication channels, uniform visual and linguistic identifiers, fast response patterns, and widely known direct communication protocols. In the Zelenskyy case, the old trust in a communications channel is an asset to be utilized in an impersonation attack. Second, the approach to communicating in a crisis should be "speed" with "evidential transparency." A denial should not just be a denial of the content but prove it is false and offer a ready alternative source of authentic content. This would help to alleviate the danger of fast correction becoming yet another battle cry. Third, governments need to enhance institutional cooperation with independent news outlets, fact-checkers and fact verification technologists. Credibility isn't something that one government or the other has, or that the media has; it's something that is jointly created, and that means there's room for official corrections to be more convincing when they are backed up by independent corroboration. Fourth, ensure that the

media organizations become institutionalized in visible verification practices. Journalists should clarify their reliance on the evidentiary basis of authentication but not be too cautious or alarmist about the use of deepfakes, as over-cautious skepticism can damage trust in genuine audiovisual evidence. Lastly, there is a need for an increase in investment in the area of deepfake literacy and ongoing readiness, not to be considered a crisis management activity. Future crisis communication should seek to achieve a calibrated level of trust – it is neither about unquestioning acceptance nor about generalised distrust, but about maintaining citizens' capacity to differentiate between credible information and fabricated communication, which is at the heart of crisis communication.

REFERENCES

1. Ahmed, S., Wang, Y., & Bee, A. W. T. "Adjusting news accuracy perceptions after deepfakes exposure: Evidence from a non-Western context." *Telematics and Informatics* 84 (2023): 102047.
2. Frandsen, F., & Johansen, W. "A brief history of crisis management and crisis communication: From organizational practice to academic discipline." *Crisis communication* 23.2 (2020): 17-58.
3. Benoit, W. L. "Image repair in crisis communication." (2016).
4. Braun, V., & Clarke, V. "Using thematic analysis in psychology." *Qualitative research in psychology* 3.2 (2006): 77-101.
5. Braun, V., & Clarke, V. "One size fits all? What counts as quality practice in (reflexive) thematic analysis?." *Qualitative research in psychology* 18.3 (2021): 328-352.
6. Cheng, Y., Li, J., Lee, J., & Liu, Y. "How is artificial intelligence shaping crisis communication? A systematic review and future research agenda." *Journal of Risk Research* (2025): 1-21.
7. Chesney, B., & Citron, D. "Deep fakes: A looming challenge for privacy, democracy, and national security." *Calif. L. Rev.* 107 (2019): 1753.
8. Coombs, W. T. "Protecting organization reputations during a crisis: The development and application of situational crisis communication theory." *Corporate reputation review* 10.3 (2007): 163-176.
9. Coombs, W. T., & Susan, Y. "Ongoing Crisis Communication: Planning, Managing, and Responding, and Robert L. Heath and H. Dan

- O'Hair, Editors Handbook of Risk and Crisis Communication Book Review." *IEEE Transactions on Professional Communication* 53.2 (2010): 174-178.
10. Coombs, W. T. "Situational crisis communication theory: Influences, provenance, evolution, and prospects." *Crisis communication* 23 (2020): 121-140.
 11. Jw, C. "Qualitative inquiry and research design." *Choosing among five traditions* (1998).
 12. Dhar, S., & Bose, I. "Victim crisis communication strategy on digital media: A study of the COVID-19 pandemic." *Decision support systems* 161 (2022): 113830.
 13. Elgueta-Ruiz, Á., Sadi, G., Gonçalves, G., & Martínez-Ortiz, J. "Meta-research on crisis communication: 24 years of research in Public Relations Review (2000–2023)." *Public Relations Review* 52.2 (2026): 102685.
 14. Ferguson, D. P., Wallace, J. D., & Chandler, R. C. "Hierarchical consistency of strategies in image repair theory: PR practitioners' perceptions of effective and preferred crisis communication strategies." *Journal of Public Relations Research* 30.5-6 (2018): 251-272.
 15. Flanagin, A. J., & Metzger, M. J. "The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information." *New media & society* 9.2 (2007): 319-342.
 16. Frandsen, F., & Johansen, W. "A brief history of crisis management and crisis communication: From organizational practice to academic discipline." *Crisis communication* 23.2 (2020): 17-58.
 17. Groh, M., Sankaranarayanan, A., Singh, N., Kim, D. Y., Lippman, A., & Picard, R. "Human detection of political speech deepfakes across transcripts, audio, and video." *Nature communications* 15.1 (2024): 7629.
 18. Hamelaers, M., Van Der Meer, T. G., & Dobber, T. "Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes." *Computers in Human Behavior* 152 (2024): 108096.
 19. Kwok, L., Lee, J., & Han, S. H. "Crisis communication on social media: what types of COVID-19 messages get the attention?." *Cornell Hospitality Quarterly* 63.4 (2022): 528-543.
 20. Masngut, N., & Mohamad, E. "Association between public opinion and Malaysian government communication strategies about the COVID-19 crisis: content analysis of image repair strategies in social media." *Journal of medical Internet research* 23.8 (2021): e28074.
 21. Metzger, M. J. "Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research." *Journal of the American society for information science and technology* 58.13 (2007): 2078-2091.
 22. Metzger, M. J., & Flanagin, A. J. "Credibility and trust of information in online environments: The use of cognitive heuristics." *Journal of pragmatics* 59 (2013): 210-220.
 23. Neyazi, T. A., Nadaf, A. H., Tan, K. E., & Schroeder, R. "Does trust in government moderate the perception towards deepfakes? Comparative perspectives from Asia on the risks of AI and misinformation for democracy." *Government Information Quarterly* 41.4 (2024): 101980.
 24. Patton, M. Q. "Qualitative research & evaluation methods: Integrating theory and practice." *Sage publications*, (2014).
 25. Rahman, B. H. "Conditional influence of media: Media credibility and opinion formation." *J. Pol. Stud.* 21 (2014): 299.
 26. Russell, K. M., & Bishop, C. O. "Understanding Ivy Lee's declaration of principles: US newspaper and magazine coverage of publicity and press agency, 1865–1904." *Public Relations Review* 35.2 (2009): 91-101.
 27. Shin, S. Y., & Lee, J. "The effect of deepfake video on news credibility and corrective influence of cost-based knowledge about deepfakes." *Digital Journalism* 10.3 (2022): 412-432.
 28. Tandoc Jr, E. C., Duffy, A., Jones-Jang, S. M., & Wen Pin, W. G. "Poisoning the information well? The impact of fake news on news media credibility." *Journal of Language and Politics* 20.5 (2021): 783-802.
 29. Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. "Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine." *Plos one* 18.10 (2023): e0291668.
 30. Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. "Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss

- deepfakes in the Russian invasion of Ukraine." *Plos one* 18.10 (2023): e0291668.
31. Vaccari, C., & Chadwick, A. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." *Social media+ society* 6.1 (2020): 2056305120903408.
 32. Vaccari, C., & Chadwick, A. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." *Social media+ society* 6.1 (2020): 2056305120903408.
 33. Cheng, Y., Wang, Y., & Kong, Y. "The state of social-mediated crisis communication research through the lens of global scholars: An updated assessment." *Public Relations Review* 48.2 (2022): 102172.
 34. Westerlund, M. "The emergence of deepfake technology: A review." *Technology innovation management review* 9.11 (2019).
 35. Yin, R. K. "Case study research and applications." Vol. 6. Thousand Oaks, CA: sage, (2018).
 36. Cheng, Y., Li, J., Lee, J., & Liu, Y. "How is artificial intelligence shaping crisis communication? A systematic review and future research agenda." *Journal of Risk Research* (2025): 1-21.
 37. Allyn, B. "Deepfake video of Zelenskyy could be tip of the iceberg in info war, experts warn." URL: <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> (2022).
 38. Hatmaker, T. "Meta takes down deepfake of Ukraine's President Zelensky surrendering." *TechCrunch*. (2022).
 39. Holroyd, M., & Olorunselu, F. "Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war." *Euro-news* (2022).
 40. Reuters. "Video shows Ukrainian President Zelenskyy surrendering. It's a deepfake." *Indian Express*. (2022).
 41. The Washington Post. "A falsified video of Ukrainian President Zelensky showed how deepfakes can be disarmed." *The article is particularly useful for the discussion of pre-bunking, rapid response, and verification.* (2022).
 42. Reuters. "What is visual verification and why does Reuters practice it?." (2026).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Smith, E. R. "Crisis Communication in the Age of Deepfakes: A Qualitative Analysis of Government and Media Response to AI-Manipulated Media" *Sarcouncil Journal of Engineering and Computer Sciences* 5.10 (2026): pp 9-23.