International Journal Publishing
INFLUENCE: International Journal of Science Review
Volume 5 No. 3, 2023
https://influence-journal.com/index.php/influence/index

Pre-Bunking Architecture for Government Public Relations Against Deepfake Disinformation

Yeti Rohayati
Universitas Langlangbuana, Bandung Indonesia
Email: yetirohayati971@gmail.com

Abstract. The rapid development of artificial intelligence (AI) has accelerated the emergence of deepfake technology, creating new challenges for government public communication. Deepfake-based disinformation has the potential to manipulate public perception, undermine institutional credibility, and weaken trust in government information. Existing government public relations approaches remain predominantly reactive, relying on fact-checking, clarification, and debunking mechanisms after false information has already circulated. This study aims to analyse the challenges posed by deepfake-based disinformation to government public communication, examine the limitations of existing public relations approaches, and formulate a pre-bunking architecture as a preventive communication model for government public relations. This research employs a qualitative design with a descriptive-analytical approach. Data were collected through a literature review of scientific journal articles, books, policy reports, and academic publications related to deepfakes, disinformation, digital governance, and strategic communication. The collected data were analysed using qualitative content analysis through classification, interpretation, comparison, and synthesis of relevant findings from previous studies. The results indicate that deepfake disinformation represents a multidimensional threat that challenges information credibility, democratic communication, and institutional legitimacy. Furthermore, existing government public relations strategies are constrained by reactive communication patterns, institutional fragmentation, limited monitoring capabilities, and insufficient public resilience-building initiatives. As a novelty, this study proposes a pre-bunking architecture consisting of four integrated components: early detection and digital monitoring, strategic pre-emptive communication, multi-stakeholder collaboration, and public digital resilience development. This architecture shifts government communication from reactive information management toward anticipatory communication governance capable of strengthening societal resilience against deepfake-based disinformation and sustaining public trust in the digital era.

Keywords: *Pre-Bunking Architecture, Government Public Relations, Deepfake Disinformation, Strategic Communication, Digital Resilience.*

A. INTRODUCTION

The rapid advancement of artificial intelligence (AI) technology has transformed the digital information ecosystem, creating both opportunities and challenges for public communication. According to Chesney & Citron (2019), deepfake technology enables the creation of highly realistic synthetic audio, video, and images capable of manipulating public perception and undermining trust in information sources. The increasing accessibility of generative AI tools has accelerated the production and dissemination of deepfake content, making it increasingly difficult for the public to distinguish between authentic and fabricated information. This phenomenon has become a significant concern for governments worldwide because deepfake-based disinformation threatens democratic processes, public trust, and social stability.

The challenge posed by deepfake disinformation is particularly relevant for government public relations, whose primary function is to maintain transparent, accurate, and trustworthy communication between government institutions and citizens. In Indonesia, the rapid growth of social media platforms has expanded the reach of information while simultaneously facilitating the spread of misleading content. Government institutions have generally adopted

reactive communication strategies focused on fact-checking, clarification, and debunking after false information has already circulated. However, the speed and sophistication of deepfake dissemination often exceed the capacity of conventional counter-disinformation mechanisms, resulting in delayed responses and diminished public confidence.

Recent studies have highlighted the growing threat of deepfake technology in political communication and public governance. Vaccari & Chadwick (2020) found that exposure to manipulated media can reduce citizens' trust in authentic information and increase uncertainty regarding political realities. Similarly, Westerlund (2019) argued that deepfakes represent a new phase of information disorder in which synthetic media can be strategically employed to influence public opinion and undermine institutional legitimacy. These findings indicate that conventional approaches to combating misinformation may be insufficient when dealing with AI-generated content that appears highly credible and emotionally persuasive.

Despite the expanding literature on disinformation and digital governance, existing research has predominantly focused on post-hoc interventions such as fact-checking systems, content moderation, media literacy campaigns, and regulatory frameworks. Studies conducted by Bennett & Livingston (2018), as well as Wardle & Derakhshan (2017), emphasize corrective and responsive measures aimed at mitigating the impact of false information after exposure occurs. While these approaches remain important, they provide limited insight into anticipatory communication models capable of preventing the acceptance of disinformation before it gains widespread traction. Consequently, the strategic role of government public relations in developing preventive communication architectures remains underexplored.

The concept of pre-bunking has recently emerged as a promising approach to countering disinformation through psychological inoculation. Roozenbeek et al. (2022) demonstrated that pre-bunking interventions can increase public resilience against manipulative information techniques by exposing audiences to weakened forms of deceptive narratives before encountering actual disinformation. Nevertheless, most pre-bunking studies concentrate on educational interventions, digital literacy tools, and platform-based initiatives. Limited scholarly attention has been devoted to how government public relations institutions can systematically integrate pre-bunking principles into their communication architecture, particularly in responding to deepfake-based disinformation threats.

This condition reveals a significant research gap. Normatively, government public relations should function as strategic communicators capable of building public resilience against emerging information threats. Empirically, however, government communication practices remain largely reactive and fragmented when confronting deepfake disinformation. Furthermore, previous studies have not comprehensively examined the institutional communication model, strategic messaging design, stakeholder collaboration mechanisms, and digital monitoring systems required to operationalize a pre-bunking architecture within government public relations frameworks. Therefore, identifying these challenges is essential to strengthen the effectiveness of public communication and safeguard public trust in the digital era.

Based on the foregoing discussion, this study offers novelty by developing the concept of a pre-bunking architecture as a government public relations communication model for countering deepfake-based disinformation. Unlike previous studies that focus on fact-checking, media literacy, or regulatory responses, this research emphasizes preventive communication strategies that proactively build societal resistance to manipulated digital content. Accordingly, the objectives of this study are: (1) to analyse the challenges posed by deepfake-based disinformation to government public communication; (2) to examine the limitations of existing government public relations approaches in addressing such threats; and (3) to formulate a pre-bunking architecture model that can be implemented by government

public relations institutions to strengthen public resilience and maintain trust in government information.

B. METHOD

This study employs a qualitative research design with a descriptive approach to analyse the communication strategies of government public relations in addressing deepfake-based disinformation through a pre-bunking architecture. Qualitative research is appropriate for exploring social phenomena, communication practices, and institutional responses in their natural contexts (Creswell & Poth, 2018). The research focuses on examining concepts, strategies, and models of preventive communication developed in the fields of public relations, digital governance, and information disorder. The research specification is descriptive-analytical, aiming to provide a comprehensive understanding of the challenges posed by deepfake disinformation and the potential role of government public relations in building public resilience against manipulated digital content.

Data were collected through a literature review of scientific journal articles, books, policy papers, reports from international organizations, and other academic publications related to deepfake technology, disinformation, pre-bunking, and government communication. According to Snyder (2019), literature review research enables the synthesis of existing knowledge to identify theoretical developments, research gaps, and emerging perspectives within a particular field. The collected data were analysed using qualitative content analysis by classifying, comparing, interpreting, and synthesizing relevant findings from previous studies. The analysis results were then used to formulate a conceptual pre-bunking architecture model that can strengthen the effectiveness of government public relations communication in preventing the spread and impact of deepfake-based disinformation.

C. RESULTS AND DISCUSSION

1. Deepfake-Based Disinformation as a Challenge to Government Public Communication

The rapid development of artificial intelligence (AI) has fundamentally transformed the contemporary information ecosystem, creating unprecedented opportunities for communication while simultaneously generating new risks associated with information manipulation. One of the most significant manifestations of this transformation is the emergence of deepfake technology, which utilizes deep learning algorithms and generative adversarial networks (GANs) to produce synthetic audio, video, and visual content that closely resembles authentic human communication. According to Roozenbeek et al. (2022) deepfakes represent a disruptive technological innovation capable of undermining public confidence in digital information because fabricated content can be made nearly indistinguishable from genuine recordings. As AI technologies become increasingly accessible, the production and dissemination of deepfake content no longer require advanced technical expertise, thereby expanding the potential for malicious actors to exploit synthetic media for political, social, and economic purposes.

The growing sophistication of deepfake technology has significantly increased the complexity of digital disinformation. Unlike traditional misinformation, which often relies on textual manipulation or misleading narratives, deepfake content incorporates visual and auditory elements that enhance perceived credibility. Vaccari & Chadwick (2020) argue that manipulated audio-visual content can exert stronger persuasive effects than text-based misinformation because audiences generally regard visual evidence as a reliable indicator of truth. Consequently, individuals exposed to deepfake videos may be more likely to accept fabricated information as authentic, particularly when the content aligns with their existing

beliefs or emotional predispositions. This characteristic makes deepfake disinformation especially dangerous within highly polarized digital environments where information is frequently consumed without adequate verification.

The challenge becomes even more significant in the context of government public communication. Government institutions rely heavily on public trust to implement policies, maintain social stability, and facilitate effective governance. Communication issued by government agencies serves not only to disseminate information but also to establish legitimacy and strengthen relationships between public institutions and citizens. However, the proliferation of deepfake content threatens these functions by creating opportunities for the manipulation of official narratives and public perceptions. Westerlund (2019) notes that deepfake technology introduces a new dimension to information disorder because fabricated content can be strategically designed to imitate political leaders, government officials, or public institutions. Such manipulation may lead citizens to question the authenticity of official communications and become increasingly sceptical toward government information.

The literature indicates that deepfake-based disinformation can influence public opinion in several ways. First, it can create false impressions regarding government policies, administrative decisions, or public statements. Deepfake videos depicting government officials making controversial remarks or announcing fictitious policies can spread rapidly across social media platforms before official clarification is issued. As a result, misinformation may shape public perception before corrective communication reaches affected audiences. Bennett & Livingston (2018) emphasize that the contemporary digital environment enables disinformation to circulate at unprecedented speed, often outpacing institutional efforts to verify and correct false information. This asymmetry between information dissemination and verification presents a substantial challenge for government communication systems.

Second, deepfake disinformation can be employed to manipulate political narratives and electoral processes. Scholars have increasingly expressed concern regarding the use of synthetic media during democratic elections and political campaigns. R. Chesney & Citron (2019) argue that deepfake technology has the potential to distort democratic deliberation by disseminating fabricated content designed to influence voter perceptions and political attitudes. In such circumstances, government public relations institutions face the difficult task of maintaining informational integrity while responding to rapidly evolving digital threats. The challenge is particularly acute in countries with large social media user populations, where digital content can reach millions of individuals within a short period of time.

Furthermore, deepfake disinformation contributes to the erosion of institutional legitimacy. Institutional legitimacy depends largely on public confidence in the credibility and reliability of government communication. When citizens encounter manipulated content purporting to originate from government officials or agencies, uncertainty regarding the authenticity of information increases. Lazer et al. (2018) explain that exposure to false information can undermine trust in authoritative sources and create confusion regarding factual reality. Over time, repeated exposure to deepfake content may weaken citizens' willingness to rely on government communication, thereby diminishing the effectiveness of public information campaigns and policy implementation efforts.

Another critical concern is the emergence of what scholars describe as the "liar's dividend." B. Chesney & Citron (2019) use this concept to explain how the existence of deepfake technology enables individuals to dismiss genuine evidence by falsely claiming that authentic recordings have been manipulated. Consequently, the challenge posed by deepfakes extends beyond fabricated content itself; it also creates opportunities for public officials, political actors, or other influential individuals to deny accountability for genuine actions and statements. This phenomenon further complicates the communication environment by

increasing public uncertainty regarding what information can be trusted. For government institutions, the liar's dividend represents a serious governance challenge because it undermines the evidentiary value of authentic digital records and weakens mechanisms of public accountability.

The threat of deepfake disinformation is amplified by the structural characteristics of social media platforms. Digital platforms prioritize engagement-driven algorithms that often reward emotionally charged and sensational content. Vosoughi et al. (2018) found that false information tends to spread faster and reach broader audiences than truthful information because it generates stronger emotional reactions and higher levels of user interaction. Deepfake content frequently exploits these dynamics by presenting dramatic, controversial, or shocking narratives that attract attention and encourage rapid sharing. Consequently, government communication units may struggle to compete with the speed and visibility of manipulated content circulating within digital networks.

The increasing prevalence of deepfakes also exposes limitations in existing public communication strategies. Traditional government communication approaches largely emphasize information dissemination, press releases, official statements, and post hoc clarification. While these mechanisms remain important, they are often insufficient to address the unique characteristics of AI-generated disinformation. Wardle & Derakhshan (2017) contend that contemporary information disorders require more proactive and preventive communication approaches rather than exclusive reliance on corrective measures. In the case of deepfake disinformation, waiting until false content has already achieved widespread circulation may significantly reduce the effectiveness of subsequent clarification efforts.

In addition to institutional challenges, deepfake disinformation poses significant societal risks by affecting citizens' ability to evaluate information critically. The increasing realism of synthetic media complicates traditional methods of source verification and media literacy. As individuals become less confident in their ability to distinguish authentic content from manipulated content, a broader crisis of informational trust may emerge. Tandoc et al. (2018) suggest that persistent exposure to misinformation can contribute to information fatigue, cynicism, and declining trust in both media organizations and public institutions. This condition may ultimately reduce civic engagement and weaken democratic participation.

Therefore, deepfake-based disinformation should not be viewed solely as a technological phenomenon but rather as a multidimensional communication challenge affecting governance, public trust, democratic stability, and information integrity. The literature consistently demonstrates that the risks associated with synthetic media extend beyond isolated incidents of misinformation and have broader implications for institutional legitimacy and societal resilience. Consequently, government public relations institutions must develop innovative communication frameworks capable of anticipating, identifying, and mitigating deepfake threats before they significantly influence public perception. Understanding the nature and implications of deepfake disinformation is therefore essential for designing effective communication strategies that can preserve public trust and strengthen the resilience of the digital public sphere.

2. Limitations of Existing Government Public Relations Approaches

The literature review reveals that government public relations practices remain predominantly reactive in addressing digital disinformation, particularly in the context of increasingly sophisticated deepfake technologies. Most existing communication strategies focus on corrective actions such as fact-checking, clarification, official statements, press releases, and counter-narratives after misleading information has already circulated within the public sphere. While these approaches are important for restoring factual accuracy, they often

fail to prevent the initial influence of false information on public perception. According to Wardle & Derakhshan (2017), the effectiveness of corrective communication is frequently limited because misinformation tends to establish cognitive impressions before factual corrections are disseminated. Once misleading narratives become embedded in public consciousness, subsequent clarification efforts may have only a partial impact on correcting misconceptions.

This challenge is closely associated with the speed of information dissemination in contemporary digital environments. Social media platforms enable content to spread across networks within minutes, often reaching millions of users before official institutions become aware of its existence. Government public relations agencies generally operate through bureaucratic procedures that require verification, approval, and coordination before information can be publicly released. Although such procedures are necessary to ensure accuracy and institutional accountability, they can significantly delay responses to rapidly evolving disinformation campaigns. Bennett & Livingston (2018) argue that digital disinformation thrives in environments characterized by information abundance and accelerated communication flows, where institutional responses frequently lag behind the speed of online content circulation. Consequently, government communication units often find themselves responding to narratives that have already gained significant public attention.

The reactive orientation of government communication is further reflected in the dominance of debunking strategies. Debunking seeks to correct false information after exposure has occurred by presenting evidence, factual clarification, or expert verification. While debunking remains an essential component of disinformation management, numerous studies suggest that corrective information does not always eliminate the effects of prior exposure. Lewandowsky et al. (2012) demonstrate that misinformation can continue influencing attitudes and decision-making even after it has been publicly corrected, a phenomenon commonly referred to as the "continued influence effect." This finding indicates that reliance on post-exposure interventions alone may be insufficient for addressing the challenges posed by deepfake-based disinformation, which often exploits emotional engagement and visual credibility to maximize persuasive impact.

Another major limitation concerns the organizational structure of government communication systems. Existing public relations frameworks frequently operate within institutional silos, where communication functions are separated from cybersecurity units, digital monitoring teams, intelligence agencies, and policy-making departments. Such fragmentation restricts the capacity of government institutions to develop integrated responses to complex information threats. Deepfake disinformation represents a multidimensional challenge involving technological, political, social, and communication dimensions simultaneously. However, many government agencies continue to address these dimensions independently rather than through coordinated governance mechanisms. According to Mergel (2013), effective digital governance requires cross-sectoral collaboration and institutional integration to address increasingly complex public communication challenges. Without such coordination, government responses are often fragmented, inconsistent, and less effective in countering rapidly emerging disinformation campaigns.

The absence of comprehensive monitoring mechanisms further contributes to institutional limitations. Many government public relations units remain focused on disseminating official information rather than systematically identifying emerging threats within digital information ecosystems. As a result, communication agencies frequently become aware of disinformation only after it has achieved substantial visibility. In contrast, contemporary information threats require proactive surveillance systems capable of detecting suspicious narratives, monitoring public sentiment, and identifying coordinated manipulation

efforts before they become widespread. Bradshaw & Howard (2018) emphasize that organized disinformation campaigns increasingly utilize automated technologies, algorithmic amplification, and networked communication strategies that can rapidly influence public discourse. Without advanced monitoring capabilities, government institutions face significant challenges in anticipating and responding to such threats effectively.

The literature also highlights limitations in institutional preparedness regarding emerging technologies. Deepfake technology evolves rapidly, continuously improving its realism and accessibility. Yet government communication frameworks often remain based on assumptions associated with traditional media environments, where information could be more easily verified and controlled. The growing availability of AI-powered content generation tools means that fabricated audio and video can now be produced at relatively low cost and distributed through multiple digital channels. B. Chesney & Citron (2019) argue that deepfakes introduce unprecedented challenges because they blur the distinction between authentic and fabricated evidence. Despite these developments, many government public relations institutions have not yet developed specialized protocols, technological capabilities, or strategic communication frameworks specifically designed to address synthetic media threats.

In addition to organizational and technological limitations, government public relations approaches often place insufficient emphasis on public engagement and citizen empowerment. Existing anti-disinformation initiatives primarily focus on delivering accurate information rather than developing public resilience against manipulation. Communication is frequently conceptualized as a one-way process in which government agencies transmit information to citizens, expecting audiences to accept official messages as authoritative. However, contemporary communication environments are characterized by participatory media cultures in which individuals actively produce, share, and interpret information. Villi (2012) note that digital media ecosystems require institutions to engage audiences as active participants rather than passive recipients. Failure to recognize this transformation limits the effectiveness of traditional communication strategies in combating disinformation.

The limited emphasis on media literacy and critical thinking further exacerbates this problem. Research indicates that citizens who possess stronger digital literacy skills are generally more capable of identifying misleading content and evaluating information credibility. Roozenbeek et al. (2022) demonstrate that psychological inoculation and media literacy interventions can significantly improve resistance to manipulative information techniques. Nevertheless, many government communication programs continue to prioritize informational campaigns over long-term resilience-building initiatives. As a result, citizens may remain vulnerable to sophisticated forms of manipulation despite having access to official information sources.

Another critical limitation concerns public trust itself. Government communication effectiveness depends substantially on the credibility of institutions in the eyes of citizens. However, trust levels vary significantly across social, political, and cultural contexts. In environments characterized by political polarization or declining institutional confidence, official corrections may be viewed with scepticism regardless of their factual accuracy. Edelman's Trust Barometer reports consistently indicate that declining trust in institutions creates fertile conditions for misinformation and conspiracy narratives. Consequently, government public relations strategies that rely exclusively on institutional authority may struggle to persuade audiences who already distrust official sources.

Furthermore, current government communication approaches often underestimate the emotional dimensions of disinformation. Research on information disorder suggests that false narratives frequently achieve influence not because they are factually convincing but because they evoke strong emotional responses such as fear, anger, outrage, or anxiety (Bakir &

McStay, 2018). Deepfake content is particularly effective in exploiting emotional reactions because audio-visual manipulation creates powerful impressions of authenticity. Nevertheless, many government communication strategies continue to prioritize factual correction without adequately addressing the emotional and psychological mechanisms that drive information sharing behaviours. This imbalance limits the ability of official communication to compete with highly engaging disinformation narratives.

Taken together, these findings demonstrate that existing government public relations approaches face significant limitations in responding to contemporary disinformation threats. Reactive communication models, institutional fragmentation, inadequate monitoring capabilities, limited technological preparedness, insufficient public engagement, and an overreliance on corrective communication collectively reduce the effectiveness of current strategies. The literature consistently suggests that combating deepfake-based disinformation requires a shift from reactive communication toward preventive and resilience-oriented approaches. Therefore, the shortcomings identified in existing government public relations practices highlight the necessity of developing an innovative communication architecture capable of anticipating disinformation threats before they influence public perception. This need provides the foundation for exploring pre-bunking as a strategic communication model that can strengthen public resilience and enhance the effectiveness of government communication in the digital era.

3. Pre-bunking Architecture as a Government Public Relations Communication Model

The findings of this study indicate that pre-bunking offers a promising and innovative framework for strengthening government public communication against the growing threat of deepfake-based disinformation. As the digital information environment becomes increasingly vulnerable to synthetic media manipulation, conventional communication approaches cantered on correction and clarification are no longer sufficient. Instead, government public relations institutions require preventive communication strategies capable of enhancing public resilience before disinformation achieves widespread influence. In this context, pre-bunking represents a strategic shift from reactive communication toward anticipatory communication. According to Roozenbeek et al. (2022), pre-bunking is based on the theory of psychological inoculation, which suggests that exposing individuals to weakened forms of manipulative techniques can strengthen their ability to recognize and resist future attempts at deception. Rather than focusing on correcting false information after exposure, pre-bunking seeks to prevent misinformation from gaining persuasive power in the first place.

The theoretical foundation of pre-bunking originates from inoculation theory developed by McGuire (1964), which argues that attitudes can be strengthened through controlled exposure to counterarguments and persuasive attacks. Contemporary scholars have adapted this concept to the digital information environment by demonstrating that individuals who understand common misinformation techniques become more resistant to deceptive content. van der Linden et al. (2020) found that pre-bunking interventions can significantly improve individuals' ability to identify manipulation tactics such as emotional exploitation, conspiracy framing, false dilemmas, and fabricated evidence. These findings are particularly relevant in the context of deepfake disinformation because synthetic media often relies on emotional persuasion and perceived authenticity to influence public opinion. Consequently, government public relations institutions can utilize pre-bunking as a proactive communication strategy to cultivate societal resilience before misinformation spreads.

Based on the synthesis of previous studies, this research proposes a pre-bunking architecture consisting of four interconnected components: early detection and digital

monitoring, strategic pre-emptive communication, multi-stakeholder collaboration, and public digital resilience development. These components collectively form an integrated communication framework designed to anticipate, mitigate, and reduce the impact of deepfake-based disinformation on public trust and institutional legitimacy.

The first component is *early detection and digital monitoring*. The literature consistently emphasizes that effective communication responses require timely identification of emerging information threats. Bradshaw & Howard (2018) argue that digital disinformation campaigns increasingly utilize coordinated networks, automated accounts, and algorithmic amplification mechanisms that enable misleading narratives to spread rapidly across online platforms. Consequently, government public relations institutions must establish continuous monitoring systems capable of tracking information flows, detecting suspicious narratives, and identifying potential deepfake content before it reaches widespread circulation.

Digital monitoring should not be limited to observing viral content after dissemination has occurred. Instead, it should involve proactive surveillance of online conversations, public sentiment trends, and emerging communication patterns. Through the integration of artificial intelligence tools, data analytics systems, and social media intelligence platforms, government communication units can identify indicators of coordinated manipulation and assess potential risks to public communication. Such capabilities allow institutions to move from a reactive posture toward an anticipatory communication model. Early detection also enhances institutional preparedness by enabling communication teams to develop response strategies before disinformation achieves significant visibility.

The second component is *strategic pre-emptive communication*, which constitutes the core mechanism of the pre-bunking architecture. Strategic pre-emptive communication involves disseminating verified information and educating audiences about potential manipulation techniques before misinformation emerges. According to Roozenbeek et al. (2022), pre-bunking interventions are most effective when they explain how manipulative communication techniques operate rather than focusing exclusively on specific false claims. This approach equips audiences with transferable cognitive skills that can be applied across different disinformation contexts.

In practical terms, government public relations institutions can implement strategic pre-emptive communication through public awareness campaigns, educational content, digital literacy initiatives, and proactive information dissemination. Citizens can be informed about the characteristics of deepfake videos, the methods commonly used to manipulate digital content, and the importance of verifying information through trusted sources. Furthermore, communication campaigns can provide examples of common manipulation strategies without exposing audiences to harmful misinformation itself. By increasing awareness of deceptive communication practices, government institutions can reduce the persuasive effectiveness of fabricated content and strengthen public confidence in authentic information sources.

The third component is *multi-stakeholder collaboration*, which recognizes that combating deepfake disinformation exceeds the capacity of individual institutions acting independently. The literature demonstrates that contemporary information threats operate across technological, political, social, and communication domains simultaneously. Consequently, effective responses require cooperation among multiple actors, including government agencies, technology companies, fact-checking organizations, academic institutions, media organizations, and civil society groups. Wardle & Derakhshan (2017) emphasize that information disorder represents a systemic challenge that necessitates collaborative governance rather than isolated institutional interventions.

Collaboration with technology platforms is particularly important because social media companies possess significant influence over information dissemination processes.

Partnerships between government communication agencies and digital platforms can facilitate the rapid identification of manipulated content and support the development of educational initiatives designed to enhance public awareness. Academic institutions contribute by conducting research on emerging disinformation trends and evaluating the effectiveness of communication interventions. Media organizations play a critical role in disseminating verified information and promoting responsible reporting practices, while civil society organizations facilitate community-level engagement and digital literacy development. Through coordinated action among these stakeholders, governments can establish a more comprehensive and adaptive communication ecosystem capable of responding effectively to deepfake threats.

The fourth component is *public digital resilience development*, which serves as the long-term objective of the pre-bunking architecture. Existing anti-disinformation strategies frequently emphasize information correction rather than resilience building. However, research indicates that sustainable protection against misinformation requires empowering citizens with the skills necessary to evaluate information critically. Guess et al. (2020) argue that digital literacy and critical thinking represent essential competencies for navigating contemporary information environments characterized by increasing complexity and manipulation.

Public digital resilience involves more than simply teaching individuals how to verify information sources. It also includes fostering awareness of cognitive biases, emotional manipulation techniques, and algorithmic influences that shape information consumption. Government public relations institutions can contribute to resilience development through educational programs targeting diverse demographic groups, including students, public servants, community leaders, and social media users. Such initiatives should emphasize critical evaluation, source verification, and responsible information-sharing practices. By strengthening citizens' cognitive resilience, governments can reduce societal vulnerability to deepfake-based disinformation and enhance the overall integrity of public communication.

The integration of these four components creates a comprehensive communication architecture that differs substantially from conventional government public relations approaches. Traditional communication models generally prioritize information dissemination and crisis response, whereas pre-bunking architecture emphasizes anticipation, preparedness, collaboration, and resilience. This shift reflects broader changes in the digital information environment, where the speed and sophistication of disinformation frequently outpace corrective interventions. Lewandowsky et al. (2017) observe, preventing misinformation from influencing public perceptions is often more effective than attempting to reverse its effects after exposure has occurred.

Moreover, the proposed architecture aligns with contemporary principles of strategic communication and digital governance. Effective governance increasingly depends on the ability of institutions to anticipate risks, engage stakeholders, and cultivate public trust within complex information ecosystems. Deepfake disinformation threatens these objectives by undermining confidence in authentic information and weakening institutional legitimacy. Therefore, integrating pre-bunking principles into government public relations strategies can contribute not only to disinformation prevention but also to broader goals of democratic governance, transparency, and citizen engagement.

Overall, the findings suggest that pre-bunking architecture provides a viable communication model for addressing the challenges posed by deepfake-based disinformation. By combining early detection, strategic pre-emptive communication, multi-stakeholder collaboration, and public resilience development, government public relations institutions can transition from reactive information management toward proactive communication governance. Such a transformation is essential in contemporary digital environments where information manipulation technologies continue to evolve rapidly. Consequently, the adoption

of pre-bunking architecture can strengthen public trust, improve institutional credibility, and enhance societal resilience against emerging forms of digital disinformation.

D. CONCLUSION

This study concludes that deepfake-based disinformation poses a significant challenge to government public communication because it undermines information credibility, weakens institutional trust, and accelerates the spread of misleading narratives within the digital public sphere. The findings reveal that existing government public relations approaches remain predominantly reactive, fragmented, and dependent on post-exposure interventions such as fact-checking and debunking, which are increasingly inadequate in responding to the speed and sophistication of synthetic media manipulation. As a novelty, this study proposes a pre-bunking architecture as a preventive government communication model consisting of four integrated components: early detection and digital monitoring, strategic pre-emptive communication, multi-stakeholder collaboration, and public digital resilience development. Through this architecture, government public relations can shift from reactive information management toward anticipatory communication governance that strengthens public resilience against deepfake-based disinformation, enhances institutional credibility, and sustains public trust in the digital era.

REFERENCES

Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions: Problems, causes, solutions. *Digital Journalism*, *6*(2). https://doi.org/10.1080/21670811.2017.1345645

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2). https://doi.org/10.1177/0267323118760317

Bradshaw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The computational propaganda project*, *1*, 1-26.

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*(6), 1753. https://doi.org/10.15779/Z38RV0D15J

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: the coming age of post-truth geopolitics. *Foreign Affairs*, *98*(1).

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publications.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27). https://doi.org/10.1073/pnas.1920498117

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380). https://doi.org/10.1126/science.aao2998

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, *6*(4). https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing.

Psychological Science in the Public Interest, Supplement, *13*(3). https://doi.org/10.1177/1529100612451018

McGuire, W. J. (1964). Inducing resistance to persuasion. Some contemporary approaches. *CC Haaland and WO Kaelber (Eds.), Self and Society. An Anthology of Readings, Lexington, Mass. (Ginn Custom Publishing) 1981, pp. 192-230.*

Mergel, I. (2013). Social media adoption and resulting tactics in the U.S. federal government. *Government Information Quarterly*, *30*(2). https://doi.org/10.1016/j.giq.2012.12.004

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34). https://doi.org/10.1126/sciadv.abo6254

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*. https://doi.org/10.1016/j.jbusres.2019.07.039

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, *6*(2). https://doi.org/10.1080/21670811.2017.1360143

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, *6*(1). https://doi.org/10.1177/2056305120903408

van Der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in psychology*, *11*, 566790. https://doi.org/10.3389/fpsyg.2020.566790

Villi, M. (2012). Social curation in audience communities: UDC (user-distributed content) in the networked media ecosystem'. *Participations: Journal of Audience and Reception Studies*, *9*(2), 614-632.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, *359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1-107). Strasbourg: Council of Europe.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11). https://doi.org/10.22215/TIMREVIEW/1282