The Research Probe
Volume 6 Issue 1 June 2026
DOI: https://doi.org/10.53378/trp.227

Seeing is deceiving: A content analysis of AI-generated political misinformation and its influence on trust in government officials

[1]Justine S. Leomo, [2]Fem Jhullian C. Reyes, [2]Princess Tiffany B. Telos, [2]Siou-Sian Chen & [2]Annika Jordan F. Dela Cruz

Abstract

Artificial Intelligence (AI) has delivered unprecedented opportunities and also spawned new societal hazards in today's rapidly evolving technological context. One of the most concerning threats involves the increasing prevalence of AI-generated political misinformation, which undermines democratic dialogue and public faith. This qualitative content analysis studied how this misinformation is created and shared in TikTok through five purposively selected videos, and audience comments on the videos. A two-cycle coding process, descriptive coding, and thematic analysis were used to conduct the qualitative content analysis. Inter-coder reliability was determined through Cohen's Kappa ($\kappa = .82$), thus showing strong agreement, and reflexivity and anonymizing were used for rigor and ethics-related issues. The study revealed five interrelated themes that illustrate the layered strategies of misinformation: polarized narratives, which reduce politics to moral binaries of heroes and villains; engineered authenticity, which uses realistic visuals and synthetic voices to mimic credibility; linguistic persuasion, which employs rhetorical and emotional cues to foster skepticism; emotional amplification, which manipulates affective responses such as outrage or admiration; and credibility hybridization, which merges real and fabricated elements to blur truth and fiction. These layers form a credibility stack, a multimodal structure combining narrative, visual, linguistic, and emotional manipulation to enhance believability. AI-generated misinformation is more persuasive and difficult to detect than traditional propaganda, hence, this study calls for strengthened media literacy initiatives and policy safeguards to mitigate its influence on democratic discourse.

Keywords: *AI-generated political contents, TikTok, political communication, content analysis, media literacy*

Article History:

Received: *February 14, 2026* ***Revised***: *May 2, 2026*
Accepted: *May 22, 2026* ***Published online***: *June 15, 2026*

Suggested Citation:
Leomo, J.S., Reyes, F.J.C., Telos, P.T.B., Chen, S. & Dela Cruz, A.J.F. (2026). Seeing is deceiving: A content analysis of AI-generated political misinformation and its influence on trust in government officials. *The Research Probe*, 6(1), 204-223. https://doi.org/10.53378/trp.227

About the authors:
[1]Corresponding author. Science, Technology, Engineering, and Mathematics student. Columban College, Inc. email: leomojustine@gmail.com
[2]Science, Technology, Engineering, and Mathematics student. Columban College, Inc.

1. Introduction

In an era defined by the rapid advancement of Artificial Intelligence (AI), many aspects of modern life were transformed, including communication, media, and politics. While AI technologies have enabled innovation and efficiency, they have also been weaponized to spread misleading content. Misinformation is not new, but AI tools have made it easier to create and share it quickly in more convincing ways (Bontridder & Poullet, 2021). This issue is particularly concerning for younger audiences who are growing up in the digital age, in which they are exposed daily to content on social media, including material that is misleading or entirely fabricated. (Pérez-Escoda et al., 2021).

One of the most concerning trends is the rise of AI-generated political misinformation videos on TikTok, which leverages deepfake technologies, voice synthesis, and automated video production to mimic authentic political speeches, interviews, or campaigns. Unlike traditional misinformation, these AI-powered outputs are harder to detect, more persuasive in presentation, and can spread rapidly across social media platforms (Surjatmodjo et al., 2024; Romanishyn et al., 2025). As Lan and Tung (2024) observed, there is a widely recognized presence of fake news on TikTok, along with differing levels of trust in the platform's content. As a result, this phenomenon poses significant challenges to democratic discourse, undermines public trust, and weakens the critical evaluation skills of audiences, potentially shaping public opinion in subtle but dangerous ways.

Despite growing concerns about AI-driven disinformation, existing studies have primarily focused on audience reaction, fact-checking strategies, or traditional "fake news" campaigns (i.e., Banlasan & Reytana, 2025; Estrellado et al., 2025; Tan et al., 2025; Kim et al., 2022; Alshwayyat & Vázquez-Herrero, 2025; Li & Chang, 2022; Berger et al., 2025). Researchers have paid far less attention to the content itself, particularly how AI-generated political misinformation employs framing strategies and narrative techniques that influence perception to people, trigger emotional responses, and make disinformation more persuasive and difficult to resist (Wang et al., 2024; Groh et al., 2022). While some qualitative studies have examined misinformation discourse, they tend to rely on interviews, surveys, or thematic content analysis, where there is limited application of more targeted approaches (Lan & Tung, 2024). This gap restricts a deeper understanding of how persuasive elements are structurally embedded in AI-generated political content. As AI technologies continue to enhance the

realism and emotional impact of misinformation, especially on platforms like TikTok, this issue is increasingly significant.

This study addresses a significant research gap by examining how AI-generated political misinformation videos on TikTok employ framing strategies and narrative techniques to shape political perception and emotional engagement. This study shifts the focus toward understanding the persuasive architecture of misinformation by identifying the dominant frames, storytelling patterns, and persuasive elements embedded in AI-produced political videos, and analyzing how these features contribute to their perceived credibility and influence. The study is motivated by the increasing visibility of manipulated political videos circulating on social media, particularly during elections and other politically sensitive periods. The emergence of generative AI technologies has transformed the landscape of political communication and enabled the production of highly realistic and emotionally compelling content that is often difficult for ordinary users to distinguish from authentic material. As misleading political narratives gain traction on algorithm-driven platforms such as TikTok, there is a growing need to investigate not only how misinformation spreads but also how it is strategically constructed to persuade audiences.

The importance of this research lies in the growing reliance of younger audiences on TikTok as a source of political information. The platform's short-form, visually engaging, and algorithmically amplified content creates conditions in which AI-generated misinformation can proliferate rapidly with limited scrutiny. By examining the framing and narrative mechanisms embedded in such content, this study contributes to a deeper understanding of how framing, agenda-setting, and persuasion operate in the era of artificial intelligence. The findings are expected to offer both theoretical and practical contributions. Theoretically, the study extends existing scholarship on political communication by exploring how AI technologies reshape traditional persuasive processes. Practically, it provides insights for policymakers, educators, media practitioners, and regulators seeking to strengthen media literacy initiatives and develop effective strategies for addressing AI-generated political misinformation. Ultimately, the study underscores the urgency of safeguarding democratic discourse by improving society's capacity to recognize and respond to increasingly sophisticated forms of digital manipulation.

2. Literature Review

2.1. The Role of AI in Transforming Media and Political Communication

The rapid advancement of generative AI brings benefits, yet concurrently introduces substantial risks, particularly in the form of increased misinformation, involving the alteration and manipulation of public perceptions. This global phenomenon challenges societal efforts to uphold information integrity and cultivate media literacy. Scholars emphasize the urgency of understanding AI's influence on public knowledge, as the volume and realism of AI-generated content have significantly outpaced traditional verification mechanisms (Chu-Ke & Dong, 2024; López-Borrull & Lopezosa, 2025). These developments are reshaping political communication, where the rapid and unchecked spread of false information undermines democratic discourse.

Global studies emphasize the significant effects of AI misinformation and disinformation. Chu-Ke and Dong (2024) contend that generative AI complicates media literacy by producing credible yet misleading content, especially in politics, thus requiring integrated media and political literacy frameworks. Likewise, López-Borrull and Lopezosa (2025) reveal that AI amplifies the speed and perceived authenticity of disinformation and exacerbate its harmful effects on democracy and underscore the need for regulatory scrutiny and heightened civic awareness. Jaidka et al. (2025) outline the lifecycle of AI-generated misinformation, particularly concerning manipulated political imagery that harms public figures' reputations, and advocate for AI-specific regulations to lessen their influence on democratic processes and media entities.

Southeast Asian nations, including the Philippines, are notably susceptible to AI-driven misinformation. According to Putra (2024), structural challenges within ASEAN, such as inadequate digital literacy, reliance on social media, and disjointed regulatory frameworks, assert the necessity for coordinated regional strategies for effective governance of AI-mediated communication. Schipper (2025) assesses the Philippine electoral landscape and indicates that AI has intensified the creation and spread of disinformation, thus jeopardizing democratic integrity. Although AI tools may improve detection and monitoring, inadequate legal structures and significant social media dependence render voters vulnerable to manipulation.

Despite an expanding body of research, notable gaps remain. Globally, investigations frequently overlook platform-specific dynamics, such as TikTok's brief video content, and rarely explore micro-level audience reception in political communication (Chu-Ke & Dong,

2024; López-Borrull & Lopezosa, 2025). Regionally, while policy analyses promote regulatory and literacy initiatives, they lack empirical evidence on voter interpretation and reactions to AI-generated political content (Putra, 2024; Schipper, 2025). Addressing these gaps is vital for enhancing theoretical understanding and developing practical countermeasures against AI-driven misinformation.

2.2. AI-Driven Political Misinformation in the Age of Social Media

Among today's generation, social media became the dominant source of information. Pérez-Escoda et al. (2021) found that 80.4% of respondents rely on digital platforms for news, with social media serving as the main gateway. Despite advancements in news distribution and fact-checking tools, misinformation persists online, influencing public opinion and diminishing trust in politicians, institutions, and media. Platforms like TikTok, where news and entertainment blend seamlessly, complicate the ability of younger users to distinguish reliable information from deceptive content (Lan & Tung, 2024). Vaccari and Chadwick (2020) found that deepfakes on social media foster confusion, doubt, and declining trust, which undermines media credibility. In political contexts, such content effectively fosters distrust toward political actors and democratic institutions. Furthermore, Pérez-Escoda et al. (2021) highlighted a generational divide: younger individuals heavily rely on social media, whereas older groups place greater trust in traditional sources, like radio and print journalism. Misinformation circulates more extensively on social media, deeming the traditional outlets as more reliable.

Disinformation campaigns involve individuals with various motivations and roles. According to Lauengco (2024), there is an increasing use of deepfakes in Philippine politics such as the misleading TikTok news clip about a presidential candidate. Social media content creators are now employing financial resources and advanced skills for deceptive political purposes. Disinformation arises from organized labor, driven by income, marketing, and moral reasoning rather than malicious intent (Ong & Cabañes, 2019). These activities mirror marketing strategies, identifying audiences, and choosing platforms. While global and ASEAN studies acknowledge the ongoing nature of disinformation and advocate for media literacy, research on youth interaction with TikTok, particularly concerning deepfakes and AI-generated content, remains limited.

2.3. Framing Strategies, Narrative Techniques, and Their Democratic Implications

Misinformation and digital discourse are influenced by framing and narration that impact perceptions and democratic engagement. Constructive discussions on misinformation are hindered by strategic emotional and value-based framing. Such strategies favor emotional responses over rational reasoning in digital contexts (Scheibenzuber et al., 2023). Similarly, narrative framing increases the credibility of misinformation, making it more shareable than straightforward facts (Greer et al., 2022). Brooks (2021) further illustrates that solution framing regarding digital threats affects the proposed implementations.

Framing strategies are employed to manipulate perceptions, particularly in political discourse. This is exemplified in the ASEAN region, particularly Malaysia, where deepfakes and AI-driven misinformation threaten politicians through election manipulation and fraudulent representations (Sabri, 2024). Such tactics blur the line between persuasion and manipulation, directly impacting democracy by fostering public confusion and eroding trust in leaders. Additionally, Philippine studies reveal that disinformation narratives reframe contested histories, obscure political myths and reshape corruption narratives into legitimacy, thus altering public understanding (Mendoza et al., 2023).

Despite insights into misinformation and framing strategies, significant gaps persist. Firstly, global studies often neglect to connect their findings to political misinformation or long-term democratic effects (Greer et al., 2022; Scheibenzuber et al., 2023). Secondly, empirical studies on audience reception and effective countermeasures in social media, especially on TikTok, are scarce (Brooks, 2021). Lastly, regional research rarely explores how framing strategies affect public perceptions and trust in government (Mendoza et al., 2023; Sabri, 2024). These gaps highlight the necessity to investigate the integration of framing strategies and narrative techniques in AI-generated content on widely utilized platforms like TikTok to enhance understanding of their influence on public perceptions and governmental trust.

2.4. Theoretical framework

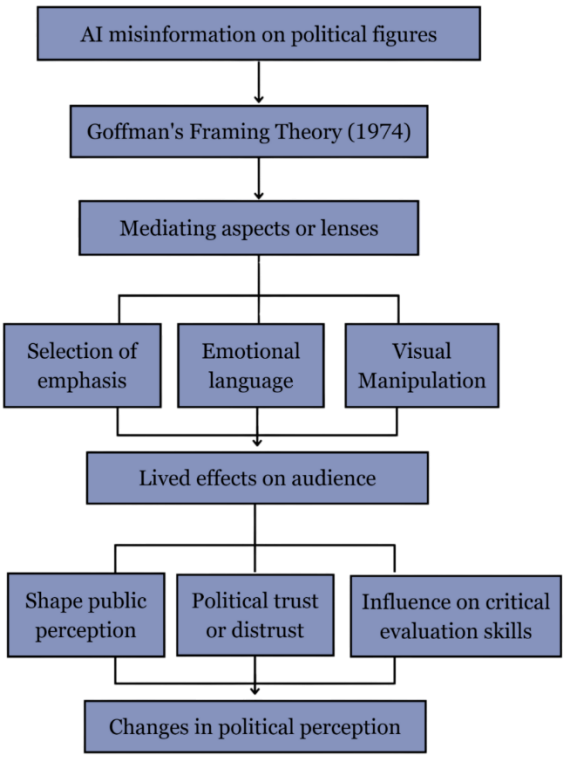
Grounded by Erving Goffman's Framing Theory (1974), this study adopts an analytical lens that moves beyond the surface description of events to interrogate the deeper processes through which meanings are actively constructed, negotiated, and contested. There are two major types of interpretive frames: the natural frame, which contextualizes events in the

unyielding realm of physical and non-social causes and the social frame, which explains events as a result of human motives, purposes and purposeful actions. The natural and social frames provide a powerful theoretical anchor that enable this study to investigate not only what happens but how individuals and groups dynamically define, justify, and reframe their lived experiences while fitting into larger social environments.

The theory explains how media and communication shape public perception by presenting information in a certain frame or context. Therefore, framing influences communication and perception due to how the information is presented, further affecting individuals' reaction and interpretation as people use these frameworks continuously in daily life to interpret events.

Figure 1

Framing theory (Goffman, 1974).



This study examines how AI is utilized to disseminate inaccurate or misleading information about politicians through the lens of Erving Goffman Framing Theory (1974). The theory posits that the manner in which information is presented influences how audiences interpret and understand social and political realities. In the context of AI-generated

misinformation, framing serves as a mechanism through which manipulated content shapes public perceptions of political figures and events.

The framing process in this study is represented through three key dimensions. First, selection of emphasis refers to the strategic highlighting of specific information, issues, or perspectives to direct audience attention toward a particular interpretation. Second, emotional language involves the use of emotionally charged words and expressions designed to evoke feelings such as fear, anger, trust, or sympathy, thereby influencing audience judgment. Third, visual manipulation pertains to the use of AI technologies to alter or generate images and videos that may distort reality, mislead viewers, or provoke desired emotional responses. These framing dimensions mediate the presentation of political information and influence how audiences process and evaluate content. As viewers are exposed to framed misinformation, their perceptions of political figures and events may be shaped through changes in trust, credibility assessments, and emotional reactions. Consequently, repeated exposure to such content may affect individuals' ability to critically evaluate information and distinguish authentic content from manipulated media.

3. Methodology

3.1. Research Design

This study employed qualitative content analysis to systematically interpret and analyze textual, visual, and audio communication patterns. This method is widely used in media and communication phenomena (Krippendorff, 2019), which examines deeper meanings and patterns embedded in political messaging rather than relying solely on numerical or statistical relationships. By focusing on interpretive analysis, the researchers identified recurring narratives, communication strategies, and authenticity markers that shape public perception. This framework was selected because AI-generated misinformation involves complex communication structures that require contextual and thematic examination rather than purely quantitative measurement. Furthermore, qualitative content analysis allowed the researchers to analyze both manifest content, which includes observable and direct messages, and latent content, which involves underlying meanings and implications within political misinformation. This design supports systematic categorization while maintaining interpretive depth, which is suitable for analyzing evolving digital political discourse.

3.2. Corpus of the Study

The corpus of the study consisted of five purposively selected TikTok videos and their corresponding audience comment sections. Instead of recruiting human respondents through interviews or surveys, the study treated digital content creators and audience interactions as units of analysis. This reflects modern research approaches in digital communication studies, where social media engagement is examined to understand public discourse and opinion formation.

The videos were selected using purposive sampling, a non-probability sampling technique that allows researchers to intentionally select samples that are highly relevant to the research objectives (Palinkas et al., 2015). The videos included in the dataset met the following criteria: (1) the content contained political messaging with identifiable AI-generated elements; (2) the videos demonstrated high engagement levels, including views, shares, and comment activity; and (3) the videos represented recent misinformation trends circulating on the platform.

3.3. Instrumentation and Data Gathering Process

The primary research instrument used in this study was a structured content analysis coding guide developed by the researchers. The coding guide served as a systematic framework for organizing and categorizing data extracted from the selected videos and audience interactions. The instrument included categories such as framing techniques, emotional tone, linguistic patterns, visual symbolism, and authenticity indicators associated with AI-generated political misinformation.

The data gathering process involved several stages. First, researchers identified and screened videos that met the inclusion criteria. Second, each selected video was carefully examined and documented. The researchers transcribed spoken dialogues, described visual components, and compiled audience comments to capture both creator intent and audience interpretation. This dual-layer analysis approach is consistent with qualitative content analysis practices that examine communication from both production and reception perspectives (Graneheim & Lundman, 2004).

To maintain consistency and reliability, data were organized into structured transcripts and coding sheets. The researchers conducted repeated viewing and reading sessions to ensure

comprehensive data familiarization. This iterative process strengthened the credibility of interpretations and minimized the risk of overlooking relevant communication patterns.

3.4. Data Analysis

Data analysis was conducted using a two-cycle coding procedure. The first cycle utilized descriptive coding, which focused on identifying and labeling observable features present in the videos and comment sections. This stage involved categorizing explicit elements such as political themes, misinformation strategies, emotional appeals, and engagement patterns. The second cycle involved thematic analysis, which aimed to identify deeper patterns and underlying meanings across the dataset. Thematic analysis enabled researchers to cluster related codes into broader conceptual themes that explain how misinformation narratives are constructed and how audiences respond to them (Braun & Clarke, 2006).

To ensure analytical rigor, coding was conducted collaboratively by two independent researchers. Inter-coder reliability was measured using Cohen's Kappa coefficient, resulting in a value of $\kappa = .82$, which indicates strong agreement between coders and demonstrates consistency in data interpretation (Cohen, 1960; Lombard et al., 2006). Discrepancies were resolved through discussion and consensus, and the finalized codes and themes were retained only after agreement was reached. This process served to ratify the findings by confirming that the identified patterns were consistently observed across coders, thereby strengthening the reliability and credibility of the results.

Additionally, researchers maintained reflective journals throughout the analysis process. These journals documented assumptions, emerging insights, and potential biases, supporting reflexivity and transparency in qualitative research (Ortlipp, 2008). Regular collaborative discussions were also conducted to resolve coding discrepancies and refine thematic categories. Although qualitative content analysis emphasizes interpretive analysis, the structured coding guide ensured systematic categorization and analytical consistency across all data sources.

3.5. Research Ethics

Ethical considerations were strictly observed throughout the research process to ensure responsible and respectful handling of digital data. The study analyzed only publicly accessible content, which eliminated the requirement for formal informed consent from individual users.

However, ethical standards were maintained by protecting participant anonymity and minimizing potential risks. Usernames and identifiable personal information were excluded, paraphrased, or anonymized to preserve confidentiality. All data were handled and stored in compliance with the Data Privacy Act of 2012 (RA 10173), ensuring proper data protection and responsible use of digital information. The study also adhered to platform guidelines by complying with TikTok's terms of service. Researchers refrained from direct interaction with users, avoided content manipulation, and maintained a non-intrusive observational approach. Furthermore, AI-generated content was analyzed using a critical yet responsible perspective, ensuring that findings were presented objectively without amplifying misinformation narratives.

4. Findings and Discussion

This study aimed to examine how AI-assisted political misinformation on TikTok is strategically constructed to shape public perception, erode trust in government officials, and influence audiences' ability to critically evaluate political information. The dataset consisted of five purposively selected TikTok videos and their corresponding comment threads. Data analysis was conducted through two coding cycles: descriptive coding was initially employed to identify the manifest characteristics and surface-level features of the content, followed by thematic analysis to uncover underlying patterns, meanings, and persuasive mechanisms.

The analysis yielded five major themes that illustrate the layered process through which AI-assisted political misinformation is constructed and presented. These themes form what this study conceptualizes as a "credibility stack," a set of interconnected elements that work together to enhance the perceived authenticity, persuasiveness, and influence of misleading political content. The succeeding sections present each theme by examining both its manifest and latent dimensions, while situating the findings within the framework of Framing Theory and relevant contemporary literature on misinformation, political communication, and digital media.

4.1. Polarized Narratives

Polarized narratives refer to the framing of political actors in rigid moral binaries, positioning them as either corrupt villains or virtuous saviors. In the dataset, this was evident in AI-generated captions such as "Exposed at Last" and "the only leader that actually cares

about us," as well as comment sections that echoed absolute trust or distrust. These observed binaries reflect how framing operates as a meaning-making mechanism in the dataset. Goffman's (1974) framing theory explains this pattern by showing that frames actively structure interpretation by assigning moral significance to events.

In this study, the use of "savior" versus "villain" labels functions as a framing device that pre-defines political actors before audiences evaluate evidence, which explains why commenters reproduce similarly absolute judgments. Ong and Cabañes (2019) directly support this mechanism by showing that disinformation strategies in the Philippines rely on moral polarization to mobilize engagement. This aligns with the dataset because the same moral extremes appear in both AI-generated content and audience responses and indicate that polarization is not only presented but reinforced through participatory amplification.

Mendoza et al. (2023) further explain this mechanism by demonstrating that cultural myths such as "Tallano Gold" persist through narrative structures of betrayal and redemption. This directly explains the dataset's polarized framing, where political meaning is structured through similar storytelling logic and allows audiences to interpret political reality through culturally familiar moral binaries. Thus, polarized narratives in this study operate through framing systems that convert political evaluation into moral alignment, which explains the observed erosion of institutional trust and the strengthening of personality-based political loyalty.

4.2. Engineered Authenticity

Engineered authenticity refers to the construction of realism in AI-generated political content through visual, auditory, and stylistic simulation of legitimate media. In the dataset, this was observed through deepfake visuals, AI-generated voiceovers, news-style formatting, and fabricated broadcast elements such as tickers and "live" indicators. This pattern reflects how credibility is produced through aesthetic cues rather than factual verification. Vaccari and Chadwick (2020) explain this mechanism by showing that deepfakes blur the boundary between real and fabricated journalism by replicating professional visual cues. This directly explains why audiences in the dataset interpreted high production quality as truth, as seen in comments like "This sounds like an actual news segment."

According to Schipper (2025), disinformation in the Philippines is intentionally designed for persuasive impact through polished presentation. This aligns with the dataset

because visual professionalism consistently functions as a credibility signal, even when content is fabricated. Through Goffman's (1974) framing theory, engineered authenticity operates by converting constructed content into "naturalized" frames, where aesthetic realism masks manipulation. This explains why audiences interpret AI-generated content as objective reporting rather than constructed narratives; generative AI ecosystems rely on surface-level credibility markers that replace source verification (López-Borrull & Lopezosa, 2025). This directly explains the observed behavior in the dataset where audiences prioritized appearance over evidence. Therefore, engineered authenticity functions as a design-based credibility system, where the appearance of professionalism becomes the primary driver of perceived truth.

4.3. Linguistic Persuasion

Linguistic persuasion refers to the strategic use of language to construct credibility, emotional direction, and interpretive meaning in political misinformation. In the dataset, this appeared through certainty markers ("proof beyond doubt"), rhetorical questions ("How long will we tolerate betrayal?"), and emotionally charged slogans that structured interpretation. These linguistic patterns are not merely stylistic but function as interpretive guides. Greer et al. (2022) explain that narrative elaboration increases persuasion by activating emotional engagement and mental imagery. This directly explains why the slogans and rhetorical structures in the dataset were repeatedly adopted and paraphrased by commenters.

Fake news framing combines emotional and argumentative language to sustain echo chambers of belief (Scheibenzuber et al., 2023). This aligns with the dataset, where repeated linguistic patterns reinforced shared suspicion toward political institutions across both videos and comments. From a framing theory perspective, linguistic persuasion functions as a "frame-through-language" mechanism, where repeated linguistic structures guide interpretation toward predefined meanings such as betrayal, exposure, and corruption. This explains why audiences co-construct misinformation narratives by reproducing the same linguistic frames, embedding them into everyday political discourse. Thus, linguistic persuasion operates not only as a transmission tool but as a cognitive structuring mechanism that shapes how political reality is interpreted.

4.4. Emotional Amplification

Emotional amplification refers to the intensification of affective responses such as anger, fear, admiration, and hope to enhance engagement and belief formation in political misinformation. In the dataset, emotional extremes were clearly visible through AI-generated content that alternated between outrage-driven accusations and idealized portrayals of political figures. These emotional cues were directly reflected in audience comments such as “We were betrayed once again” and “Finally, someone strong enough to lead us,” indicating that emotional framing shaped interpretation and engagement. This emotionally driven misinformation reduces institutional trust by prioritizing affect over verification (Pérez-Escoda et al., 2021) and directly explains why emotional content in the dataset consistently displaced evidence-based evaluation.

According to Lan and Tung (2024), younger social media users are particularly vulnerable to emotionally amplified misinformation due to algorithmic exposure and limited verification behavior. This aligns with the TikTok-based dataset, where emotional intensity was a primary driver of engagement. Within framing theory, emotional cues operate as interpretive triggers that structure political meaning through affective categories such as betrayal, hope, and fear, which also explains why emotional amplification in the dataset simultaneously weakened systemic trust while strengthening selective political loyalty. Thus, emotional amplification functions as a mechanism that reorganizes political perception by replacing factual evaluation with affective interpretation.

4.5. Credibility Hybridization

Credibility hybridization refers to the blending of authentic and AI-generated elements within a single narrative to produce hybrid forms of misinformation that appear partially real. In the dataset, this was observed in videos combining real rally footage, speeches, or press events with AI-generated voiceovers, altered subtitles, and manipulated visuals. This blending creates credibility precisely because it embeds falsehood within familiar reality. Chu-Ke and Dong (2024) explain this mechanism as a hybridization effect, where partial truth increases the perceived plausibility of fabricated content. This directly explains why viewers interpreted hybrid content as authentic, especially when it included recognizable real-world contexts.

According to López-Borrull and Lopezosa (2025), generative AI systems create environments where truth and fabrication coexist which makes verification increasingly

ineffective. This aligns with the dataset, where audiences struggled to distinguish manipulated narratives from genuine events. From Goffman's (1974) framing theory, hybridized content produces synthetic frames that imitate reality while embedding distortion within familiar contexts. This explains why audiences rely on personal experience as validation, as seen in comments like "I was at that rally, so this must be real." Credibility hybridization demonstrates a shift in misinformation logic where persuasion no longer depends on complete falsity, but on strategically blending truth and fabrication to exploit experiential trust.

5. Discussion

The study findings indicate that AI-generated political misinformation on TikTok is organized through a hierarchical structure of credibility. Polarized narratives reduce politics to binary roles, engineered authenticity creates a simulated reality through design, linguistic persuasion heightens paranoia with certainty markers, emotional amplification fosters belief in evidence and hybridized credibility blends truth and fiction to increase misinformation prevalence. Each layer functions in concert. Narratives provide the storyline, visuals and sounds confer authority, language incites paranoia, emotion reinforces belief, and hybridization ensures plausibility. It is not mere misinformation but a multimodal ecology that reshapes political trust and evaluative scrutiny.

Theoretically, this research elaborates on Goffman's (1974) framing theory that AI facilitates the construction of a synthetic frame wherein truth and fiction are intertwined to create new political realities. Practically, these results call for urgent necessity for interventions through media literacy curricula to educate audiences on identifying hybridized misinformation cues, along with policies mandating disclosure of AI-generated content (Jaidka et al., 2025). Regionally, the findings support Putra's (2024) appeal for ASEAN to address disinformation as a systemic threat, while Schipper (2025) contended that "disinformation by design" jeopardizes electoral integrity.

This study reveals that AI-generated TikTok misinformation constitutes an intricate choreography of multimodal strategies aimed at undermining institutional trust, consolidating partisan loyalty, and obstructing evaluative criticism. Future research must explore the strategic mechanics across platforms, historical variations, trust severity, and the assessment of progressive policy alternatives like provenance labels, algorithmic friction, and participatory literacy curricula.

6. Conclusion

This study demonstrate that AI-generated political misinformation on TikTok operates as an integrated system of perception shaping rather than isolated deceptive content. Across the identified themes, misinformation was shown to function through multimodal strategies that combine synthetic audio-visual manipulation, narrative framing, emotional intensification, and hybridized representations of reality. These mechanisms work together to restructure how political information is interpreted, weakening institutional trust while simultaneously shaping belief formation and critical evaluation processes. The results indicate that AI-mediated misinformation does not merely distort political facts but actively reorganizes the conditions under which political meaning is constructed and understood in digital environments.

The analysis identifies three overlapping and reinforcing themes in AI-generated misinformation. First is the manipulation of perception through AI-generated audio recordings that convincingly simulated political figures, lending false statements an aura of authenticity. Second is the use of satirical critiques that diminish political credibility, by blurring the boundaries between parody, entertainment, and deception. While satire has always been associated with political discourse, AI-enhanced satire often circulates without contextual cues, making it difficult for the audience to discern facts from irony. Third is the fabricated scandals leveraging deepfake visuals that fused real and synthetic elements to create believable portrayals of misconduct. These patterns illustrate how AI-based misinformation operates through a sophisticated multimodal layering of narrative, visuals, language, and affect, amplifying persuasive power and accelerating the erosion of institutional trust.

Beyond descriptive analysis, the findings demonstrate that AI-generated misinformation functions as a credible stack in which fact and fiction are strategically integrated. It does not merely misrepresent reality and facts, but also actively reshapes it. This process intensifies traditional propaganda techniques by making them more immersive, emotionally resonant, and difficult to detect. In doing so, the study extends existing scholarship on framing and misinformation by showing that contemporary AI-mediated framing does not merely shape interpretation but actively constructs political reality. The research questions were therefore substantially answered, as the findings explain how AI-generated content influences belief formation, weakens trust in public institutions, and complicates citizens' capacity for critical evaluation.

The implications of these findings are both practical and immediate. From an educational standpoint, the results highlight the urgency of strengthening digital and media literacy programs that move beyond surface-level factchecking toward deeper analytical skills, such as recognizing synthetic media cues and understanding algorithmic amplification. This is essential in order to enhance citizens' ability to critically evaluate political content. Such efforts must go beyond basic fact-checking skills and require training in recognizing synthetic media, emotional manipulation, and algorithmic amplification. At the policy level, the study supports the development of regulatory frameworks that require transparency, labeling, and accountability for AI-generated political content, particularly on high-velocity platforms such as TikTok. These measures are essential to safeguarding democratic discourse without undermining legitimate political expression.

In terms of broader societal impact, this research underscores the responsibility shared by governments, technology platforms, educators, and civil society organizations in addressing AI-driven misinformation. Effective responses require trans-sectoral collaboration, continuous public education, and adaptive governance mechanisms capable of responding to rapidly evolving technologies. The study contributes actionable insights that may inform platform moderation strategies, civic education curricula, and public policy interventions aimed at strengthening democratic trust.

Despite its contributions, this study is not without limitations. Its focus on TikTok as a single platform limits the generalizability of the findings across other social media environments with different affordances and audience dynamics. Additionally, the fast-paced evolution of AI technologies means that specific misinformation techniques may change over time. Future research should therefore examine cross-platform misinformation ecosystems, conduct longitudinal analyses on the long-term effects of AI-generated content on political trust, and explore comparative studies across different political systems and cultural contexts.

This study affirms that developing resilience against AI-generated political misinformation is an ongoing and collective challenge. While the task is complex, the findings suggest that informed policy, critical education, and sustained scholarly inquiry can strengthen democratic capacity in the face of technological manipulation. The future of democracy increasingly depends on citizens who are not only digitally connected but critically equipped to navigate AI-mediated political realities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was not supported by any funding.

AI Declaration

The authors declare the use of Artificial Intelligence (AI) in writing this paper. In particular, the authors used Grammarly in checking grammar, summarizing key points and paraphrasing ideas. The authors take full responsibility in ensuring proper review and editing of contents generated using AI.

ORCID

Justine S. Leomo – https://orcid.org/0009-0003-3929-8019

Fem Jhullian C. Reyes – https://orcid.org/0009-0003-0272-9823

Siou-Sian Chen – https://orcid.org/0009-0001-4110-1700

Princess Tiffany B. Telos – https://orcid.org/0009-0000-1369-8813

Annika Jordan F. Dela Cruz – https://orcid.org/0009-0005-2119-5294

References

Alshwayyat, H., & Vázquez-Herrero, J. (2025). Fact-checking platforms in the Middle East: A comparative study in the age of artificial intelligence. *Social Sciences, 15*(3), Article 185. https://doi.org/10.3390/socsci15030185

Banlasan, J. A. G., & Reytana, J. T. (2025). The perceived influence of propaganda in understanding civic issues. *The Research Probe, 5*(2), 87–105. https://doi.org/10.53378/trp.196

Berger, L. M., Kerkhof, A., Mindl, F., & Münster, J. (2025). Debunking "fake news" on social media: Immediate and short-term effects of fact-checking and media literacy interventions. *Journal of Public Economics, 245*, Article 105345. https://doi.org/10.1016/j.jpubeco.2025.105345

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy, 3*, e32. https://doi.org/10.1017/dap.2021.20

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brooks, C. F. (2021). Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 159–163. https://doi.org/10.1089/cyber.2020.0183

Chu-Ke, C., & Dong, Y. (2024). Misinformation and literacies in the era of generative artificial intelligence: A brief overview and a call for future research. *Emerging Media, 2*(1), 70–85. https://doi.org/10.1177/27523543241240285

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Estrellado, C. J. P., Millar, G. B., Rey, F. I., & Arcinas, M. M. (2025). Deepfakes, of prey? Precursors and eclipses to the macrocosm Philippine politics. *International Review of Social Sciences Research, 5*(3), 26–53. https://doi.org/10.53378/irssr.353234

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harper & Row.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105–112. https://doi.org/10.1016/j.nedt.2003.10.001

Greer, J., Fitzgerald, K., & Vijaykumar, S. (2022). Narrative elaboration makes misinformation and corrective information regarding COVID-19 more believable. *BMC Research Notes, 15*, Article 235. https://doi.org/10.1186/s13104-022-06134-9

Groh, M., Sankaranarayanan, A., Kim, D., Lippman, A., & Picard, R. (2024). Human detection of political speech deepfakes across transcripts, audio, and video. *Nature Communications, 15*. https://doi.org/10.1038/s41467-024-51998-z

Jaidka, K., Chen, T., Chesterman, S., Hsu, W., Kan, M.-Y., Kankanhalli, M., Lee, M. L., Seres, G., Sim, T., Taeihagh, A., Tung, A., Xiao, X., & Yue, A. (2025). Misinformation, disinformation, and generative AI: Implications for policy and research. *Digital Government: Research and Practice, 6*(1), 1–15. https://doi.org/10.1145/3689372

Kim, H. S., Suh, Y. J., Kim, E.-M., Chong, E., Hong, H., Song, B., Ko, Y., & Choi, J. S. (2022). Fact-checking and audience engagement: A study of content analysis and audience behavioral data of fact-checking coverage from news media. *Digital Journalism, 10*(5), 781–800. https://doi.org/10.1080/21670811.2021.2006073

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications.

Lan, D. H., & Tung, T. M. (2024). Exploring fake news awareness and trust in the age of social media among university student TikTok users. *Cogent Social Sciences, 10*(1), Article 2302216. https://doi.org/10.1080/23311886.2024.2302216

Lauengco, G. (2024). AI deep fakes – technology and politics 2025. *Philippine News Agency*. https://tinyurl.com/4bp7yfmc

Li, J., & Chang, X. (2022). Combating misinformation by sharing the truth: A study on the spread of fact-checks on social media. *Information Systems Frontiers*, 1–15. Advance online publication. https://doi.org/10.1007/s10796-022-10296-z

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2006). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

López-Borrull, A., & Lopezosa, C. (2025). Mapping the impact of generative AI on disinformation: Insights from a scoping review. *Publications, 13*(3), 33. https://doi.org/10.3390/publications13030033

Mendoza, R. U., Elemia, C. K. S., Recto, J. M. M., & de Castro, B. A. B. (2023). When fake news infects political networks: Case study of the Tallano Gold myth in the Philippines. *Media Asia, 50*(4), 501–527. https://doi.org/10.1080/01296612.2023.2217607

Ong, J. C., & Cabañes, J. V. A. (2019). When disinformation studies meets production studies: Social identities and moral justifications in the political trolling industry. *International Journal of Communication, 13*, 5771–5790. https://ijoc.org/index.php/ijoc/article/view/11417

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report, 13*(4), 695–705. https://doi.org/10.46743/2160-3715/2008.1579

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Pérez-Escoda, A., Pedrero-Esteban, L. M., Rubio-Romero, J., & Jiménez-Narros, C. (2021). Fake news reaching young people on social networks: Distrust challenging media literacy. *Publications, 9*(2), 24. https://doi.org/10.3390/publications9020024

Putra, B. A. (2024). Fake news and disinformation in Southeast Asia: How should ASEAN respond? *Frontiers in Communication, 9*, Article 1380944. https://doi.org/10.3389/fcomm.2024.1380944

Romanishyn, A., Malytska, O., & Goncharuk, V. (2025). AI-driven disinformation: Policy recommendations for democratic resilience. *Frontiers in Artificial Intelligence, 8*, 1569115. https://doi.org/10.3389/frai.2025.1569115

Sabri, F. A. A. (2024). Narrative review on detonation of AI-generated deep fakes and misinformation: A threat to politicians in Malaysia. *E-Journal of Media and Society, 7*(2), 71–88. https://ir.uitm.edu.my/id/eprint/97612

Scheibenzuber, C., Neagu, L.-M., Ruseti, S., Artmann, B., Bartsch, C., Kubik, M., Dascalu, M., Trausan-Matu, S., & Nistor, N. (2023). Dialog in the echo chamber: Fake news framing predicts emotion, argumentation and dialogic social knowledge building in subsequent online discussions. *Computers in Human Behavior, 140*, Article 107587. https://doi.org/10.1016/j.chb.2022.107587

Schipper, T. (2025). Disinformation by design: Leveraging solutions to combat misinformation in the Philippines' 2025 election. *Data & Policy, 7*, Article e39. https://doi.org/10.1017/dap.2025.18

Surjatmodjo, D., Unde, A., Cangara, H., & Sonni, A. (2024). Information pandemic: A critical review of disinformation spread on social media and its implications for state resilience. *Social Sciences, 13*(8), 418. https://doi.org/10.3390/socsci13080418

Tan, W., Su, Y., Liu, X., Akmatbekova, Z., & Halilov, A. (2025). Research on the causes and coping strategies of the spread of false news in the era of social media. *News Research Journal, 47*(1), 86–110. https://doi.org/10.1177/30497841251404968

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception. *Social Media + Society, 6*(1), 1–13. https://doi.org/10.1177/2056305120903408

Wang, G., Frederick, R., Duan, J., Wong, W., Rupar, V., Li, W., & Bai, Q. (2025). Detecting misinformation through framing theory: The frame element-based model. *Journal of Computational Social Science, 8*. https://doi.org/10.1007/s42001-025-00403-w