AI Deepfakes and the Erosion of Stakeholder Trust: Ethical Dilemmas in Corporate Communication in Kenya

[*1]Elizabeth Ngele Mwakudwa & [2]Prof. Karim Omido

[1]PhD Student, Department of Business and Economics, Taita Taveta University, Kenya

[2]Supervisor, Department of Business and Economics, Taita Taveta University, Kenya

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100600162

Received: 20 May 2026; Accepted: 25 May 2026; Published: 19 June 2026

ABSTRACT

This study examined the influence of AI deepfakes on stakeholder trust in corporate communication within Kenyan organizations, with ethical dilemmas considered as a moderating factor. The study was motivated by the increasing use of synthetic media technologies and the growing risks they pose to organizational credibility, communication authenticity, and stakeholder confidence in digitally connected environments. A qualitative research design was adopted, and data were collected through semi-structured interviews, focus group discussions, and observations involving stakeholders from selected Kenyan corporations. Thematic analysis was used to identify key patterns and perceptions related to deepfake exposure, sophistication, detection difficulty, and ethical governance challenges. The findings revealed that repeated exposure to deepfakes, increasing realism of manipulated content, and verification challenges contribute significantly to stakeholder skepticism and reputational vulnerability. The study further established that ethical leadership, transparent communication practices, and effective verification mechanisms can mitigate trust erosion. The research contributes to emerging scholarship on AI governance, corporate communication, and stakeholder trust in African digital environments.

Keywords: *AI Deepfakes, Stakeholder Trust Erosion, Corporate Communication, Ethical Dilemmas, Detection Difficulty*

INTRODUCTION

Corporate communication has become a strategic organizational function that shapes stakeholder relationships, institutional legitimacy, and corporate reputation in increasingly digitalized business environments. In contemporary organizations, communication is no longer limited to information dissemination but serves as a mechanism for building transparency, accountability, and trust among investors, employees, regulators, consumers, and the public. The rapid expansion of digital platforms, artificial intelligence (AI), and social media ecosystems has transformed how organizations engage stakeholders, creating both opportunities for efficiency and significant risks associated with misinformation and digital manipulation (Cornelissen, 2020; Zerfass et al., 2022). In Kenya, where mobile penetration, internet accessibility, and social media usage continue to expand rapidly, corporate communication increasingly depends on digital channels that are vulnerable to emerging forms of synthetic media manipulation (Communications Authority of Kenya, 2024). As a result, organizations face growing pressure to preserve information authenticity and maintain stakeholder trust within a highly mediated communication environment.

One of the most disruptive developments within the digital communication landscape is the emergence of AI-generated deepfakes. Deepfakes refer to synthetic audio, video, or image content generated using deep learning algorithms, particularly generative adversarial networks (GANs), to imitate real individuals with a high degree of realism (Chesney & Citron, 2019). Although AI technologies offer legitimate applications in marketing,

virtual training, customer engagement, and automated communication, deepfake technologies have simultaneously introduced significant ethical and organizational risks (Kietzmann et al., 2020). Scholars argue that deepfakes undermine epistemic trust by making it increasingly difficult to distinguish authentic communication from manipulated content (Vaccari & Chadwick, 2020). The growing sophistication of synthetic media has therefore intensified concerns regarding misinformation, reputational manipulation, fraud, and identity deception across corporate environments (Westerlund, 2019).

Within organizational settings, stakeholder trust constitutes a critical intangible asset that influences organizational legitimacy, investment confidence, customer loyalty, and employee commitment. According to stakeholder theory, organizations depend on sustained trust-based relationships with multiple stakeholder groups to maintain operational stability and long-term performance (Freeman, 1984). However, the rise of deepfake technologies threatens these relationships by weakening confidence in digital corporate communication. AI-generated impersonations of executives, falsified announcements, fabricated financial statements, and manipulated corporate videos increasingly expose organizations to reputational crises and strategic communication failures (Citron & Chesney, 2019). Research by Westerlund (2019) further observes that deepfakes create a "liar's dividend," whereby stakeholders become uncertain not only about manipulated content but also about authentic organizational communication. Consequently, deepfakes contribute to broader crises of informational authenticity and digital trust within contemporary corporate ecosystems (Allcott & Gentzkow, 2017).

Globally, the proliferation of AI-generated synthetic media has accelerated concerns regarding cybersecurity, strategic communication, and organizational governance. Recent studies indicate that generative AI technologies are increasingly exploited for financial fraud, social engineering attacks, political disinformation, and corporate impersonation schemes (Kietzmann et al., 2020; Diakopoulos & Johnson, 2021). The World Economic Forum identifies misinformation and AI-enabled manipulation among the most significant emerging risks affecting institutions, democratic systems, and business operations in the digital economy (World Economic Forum, 2024). Deepfake-enabled voice cloning and executive impersonation attacks have already resulted in substantial financial losses for corporations across multiple jurisdictions, demonstrating the strategic vulnerability of digital corporate communication infrastructures (Europol, 2022). As synthetic media technologies become more accessible and difficult to detect, organizations face escalating challenges in maintaining communication credibility and safeguarding stakeholder confidence (Taddeo & Floridi, 2018).

The African context presents unique vulnerabilities associated with AI deepfakes due to rapid digital adoption, uneven technological regulation, and varying levels of digital literacy. Across the continent, financial institutions, fintech platforms, media organizations, and public institutions increasingly encounter AI-enabled fraud and impersonation threats (African Union, 2023). The expansion of mobile banking, digital finance, and social networking platforms has simultaneously increased opportunities for innovation and exposure to cyber-enabled misinformation campaigns (GSMA, 2023). Emerging evidence suggests that deepfake-enabled scams and biometric manipulation are becoming more prevalent within African digital ecosystems, complicating identity verification and intensifying cybersecurity concerns (Smile ID, 2025). These developments raise critical questions regarding institutional preparedness, regulatory capacity, and ethical governance within rapidly evolving digital societies (Mutsvairo & Ragnedda, 2019).

In Kenya, the risks associated with AI deepfakes are particularly significant because of the country's highly connected digital environment and expanding corporate reliance on online communication platforms. Kenyan organizations increasingly use digital media for investor relations, customer engagement, crisis communication, and strategic branding (Mare, 2020). However, the same communication infrastructure has facilitated the rapid spread of manipulated content through platforms such as WhatsApp, Facebook, Telegram, and TikTok. Recent Kenyan studies demonstrate growing concern among communication practitioners regarding the reputational and crisis management implications of AI-generated deepfakes. For instance, Karanja and Nzau (2026) found that public relations practitioners in Kenya perceive deepfakes as a major threat to strategic crisis communication because of their realism, rapid dissemination, and capacity to undermine public trust before verification occurs. The study further notes that many organizations remain inadequately prepared to detect or respond effectively

to synthetic media threats, highlighting deficiencies in technical expertise, ethical guidance, and institutional coordination (Karanja & Nzau, 2026).

Beyond technological risks, deepfakes also generate complex ethical dilemmas surrounding transparency, accountability, consent, privacy, and responsible AI governance. Iyer and Gokhale (2026) argue that the convergence of AI and synthetic deception creates tensions between technological innovation and informational authenticity, thereby destabilizing stakeholder confidence in organizational decision-making processes. In corporate communication contexts, ethical dilemmas emerge when organizations must balance rapid digital engagement with verification responsibilities, disclosure obligations, and reputational management (Floridi et al., 2018). Failure to establish clear ethical frameworks for AI governance may intensify stakeholder skepticism and weaken organizational legitimacy (Martin, 2019). Consequently, the challenge facing Kenyan corporations is not only technological but also ethical and strategic.

Statement of the Problem

The increasing adoption of artificial intelligence (AI) in corporate communication has transformed how organizations engage stakeholders through personalized messaging, automated customer interaction, and real-time digital engagement. However, alongside these advancements has emerged the growing threat of AI-generated deepfakes, which undermine the authenticity and credibility of organizational communication. Globally, deepfake incidents have increased significantly in recent years, with synthetic media increasingly being used in financial fraud, misinformation campaigns, executive impersonation, and reputational sabotage (Kietzmann et al., 2020). In Africa, rapid digitalization and high mobile penetration have intensified organizational exposure to AI-enabled manipulation and cyber deception. Kenya, which has one of the highest rates of internet and social media usage in Sub-Saharan Africa, faces heightened vulnerability due to widespread reliance on digital communication platforms such as WhatsApp, Facebook, and TikTok for corporate and public engagement (Communications Authority of Kenya, 2024). The sophistication of deepfakes, including realistic voice cloning, facial simulation, and contextual manipulation, has made detection increasingly difficult, thereby eroding stakeholder trust, intensifying skepticism, and exposing organizations to reputational and financial risks (Hussein & Özad, 2025).

Despite these emerging risks, many Kenyan corporations remain inadequately prepared to manage the ethical and strategic implications of deepfake technologies. Existing organizational responses are often constrained by limited detection capabilities, inadequate AI governance frameworks, weak regulatory implementation, and insufficient employee awareness regarding synthetic media threats (Karanja & Nzau, 2026). Consequently, organizations experience growing communication vulnerabilities characterized by misinformation, delayed crisis response, declining stakeholder confidence, and uncertainty regarding the authenticity of digital corporate content. Ethical dilemmas surrounding transparency, accountability, disclosure obligations, and responsible AI usage further complicate corporate responses to deepfake incidents. Without clear ethical guidelines and proactive communication strategies, organizations risk amplifying stakeholder distrust and weakening institutional legitimacy within Kenya's increasingly digital business environment (Floridi et al., 2018; Martin, 2019).

Although scholarly interest in AI, misinformation, and digital manipulation continues to expand, limited empirical research has specifically examined the relationship between AI deepfakes and stakeholder trust within Kenyan corporate communication contexts. Existing Kenya-focused studies primarily emphasize legal regulation, political misinformation, or cybersecurity concerns while giving limited attention to organizational trust dynamics and ethical communication challenges in corporate settings (Kigwiru, 2022; Amatika, 2022). Similarly, most African studies remain conceptual or policy-oriented, with insufficient qualitative evidence on how stakeholders interpret and respond to deepfake-driven communication threats in everyday organizational practice (Karanja & Nzau, 2026). This creates a significant contextual and conceptual gap in understanding how ethical dilemmas moderate the relationship between AI deepfakes and stakeholder trust erosion in Kenyan corporations. The present study therefore seeks to address this gap by examining the influence of AI deepfakes on stakeholder trust in Kenyan corporate communication, with particular emphasis on the moderating role of ethical dilemmas.

AIM AND OBJECTIVES

The main objective of this study is to examine the influence of AI deepfakes on the erosion of stakeholder trust in corporate communication in Kenya, with ethical dilemmas in corporate communication acting as a moderating variable. The specific objectives of the study are:

1. To determine the influence of the frequency of exposure to AI deepfakes on the erosion of stakeholder trust in Kenyan corporate communication.
2. To evaluate the effect of the sophistication level of AI deepfakes on the erosion of stakeholder trust among stakeholders of Kenyan corporations.
3. To assess the relationship between the detection difficulty of AI deepfakes and the erosion of stakeholder trust in corporate communication in Kenya.
4. To examine the moderating effect of ethical dilemmas on the relationship between AI deepfakes and the erosion of stakeholder trust in corporate communication in Kenya.

Research Questions

1. How does the frequency of exposure to AI deepfakes influence stakeholder trust in Kenyan corporate communication?
2. What is the effect of the sophistication level of AI deepfakes on stakeholder trust erosion?
3. How does detection difficulty of AI deepfakes relate to trust erosion in corporate settings?
4. How do ethical dilemmas moderate the relationship between AI deepfakes and stakeholder trust?

LITERATURE REVIEW

Theoretical Framework

This study is anchored on Cultivation Theory, the Elaboration Likelihood Model (ELM), and Signal Detection Theory, which collectively provide a multidimensional explanation of how AI deepfakes influence stakeholder trust in corporate communication. These theories are relevant because they explain the cumulative effects of media exposure, the persuasive influence of sophisticated synthetic content, and the cognitive difficulties stakeholders experience when attempting to distinguish authentic communication from manipulated information within digital environments.

Cultivation Theory, developed by George Gerbner and colleagues in 1969, posits that prolonged exposure to mediated content gradually shapes individuals' perceptions of social reality (Gerbner & Gross, 1976). The theory argues that repeated interaction with consistent media messages cultivates shared beliefs, attitudes, and behavioral expectations over time. Although initially developed in relation to television exposure, contemporary scholars have extended the theory to digital and social media ecosystems characterized by continuous information circulation and algorithmically amplified content (Morgan et al., 2015). In the context of AI deepfakes, repeated exposure to manipulated audio, video, and image content may cultivate generalized skepticism toward digital communication, particularly when stakeholders become uncertain about the authenticity of online information. Within Kenya's highly connected communication environment, where social media platforms and messaging applications such as WhatsApp and TikTok dominate information dissemination, persistent exposure to deepfakes may normalize distrust and weaken confidence in corporate communication. Cultivation Theory therefore provides an important explanatory framework for understanding how sustained exposure to synthetic media contributes to long-term erosion of stakeholder trust, reputational credibility, and organizational legitimacy.

The Elaboration Likelihood Model (ELM), advanced by Richard Petty and John Cacioppo in 1986, explains how individuals process persuasive information through two primary routes: the central route and the peripheral route (Petty & Cacioppo, 1986). The central route involves careful, analytical evaluation of information, while the peripheral route relies on superficial cues such as attractiveness, realism, authority, or emotional appeal. In digital communication environments characterized by information overload, audiences frequently rely on

peripheral cues when evaluating online content (Kitchen et al., 2014). This theory is particularly relevant to AI deepfakes because highly sophisticated synthetic media often mimic credible visual, vocal, and contextual signals that create perceptions of authenticity. Deepfakes featuring executive impersonation, realistic speech synthesis, or contextual corporate branding can therefore persuade stakeholders before critical verification occurs. In corporate communication contexts, such manipulative cues may influence investor confidence, customer perceptions, and organizational reputation by exploiting cognitive shortcuts associated with trust formation. The theory further explains why detection of deception after initial persuasion frequently results in reputational backlash, skepticism, and distrust toward subsequent organizational communication. ELM thus provides a useful framework for understanding how the sophistication and realism of AI deepfakes shape stakeholder perceptions and ethical dilemmas surrounding authenticity in Kenyan corporate communication.

Signal Detection Theory, developed by David Green and John Swets in 1966, explains how individuals distinguish meaningful signals from noise under conditions of uncertainty (Green & Swets, 1966). The theory emphasizes perceptual sensitivity, decision-making thresholds, and the probability of errors such as false positives and false negatives when identifying signals within complex environments. In the context of AI deepfakes, authentic corporate communication represents the signal, while manipulated synthetic media constitutes noise. As deepfake technologies become increasingly realistic and difficult to detect, stakeholders may struggle to accurately distinguish authentic organizational communication from deceptive content. This challenge is intensified by technological limitations, cognitive biases, limited digital literacy, and the rapid dissemination of information across social media networks (Vaccari & Chadwick, 2020). Within Kenyan corporate environments, detection difficulties may generate uncertainty, communication hesitancy, and generalized distrust among stakeholders, particularly where organizations lack robust verification systems or ethical communication frameworks. Signal Detection Theory therefore offers a valuable conceptual lens for understanding how verification challenges contribute to stakeholder skepticism and trust erosion in digital corporate communication.

EMPIRICAL REVIEW

Frequency of Exposure to AI Deepfakes and Stakeholder Trust Erosion

Existing empirical studies demonstrate that repeated exposure to AI-generated deepfakes significantly influences public perceptions of credibility, authenticity, and trust in digital communication environments. Research within media and communication studies shows that persistent interaction with manipulated media gradually normalizes skepticism toward online information and weakens confidence in institutional communication. George Gerbner's cultivation perspective has been widely applied in recent digital misinformation studies to explain how sustained exposure to deceptive content shapes long-term attitudes and beliefs. Empirical evidence by Cristina Vaccari and Andrew Chadwick found that exposure to synthetic media contributes to declining confidence in digital news and online communication systems, even when audiences are aware that manipulated content exists (Vaccari & Chadwick, 2020). Similarly, recent studies indicate that frequent encounters with deepfakes blur distinctions between authentic and manipulated information, thereby increasing uncertainty and encouraging generalized distrust toward mediated communication. Scholars increasingly argue that the danger of deepfakes lies not only in deception itself but also in the cumulative erosion of confidence in all forms of digital evidence, including legitimate organizational communication (Westerlund, 2019).

Empirical research further suggests that frequency of exposure intensifies the psychological and behavioral effects of misinformation within digitally saturated environments. Studies on deepfake dissemination reveal that repeated encounters with synthetic media across social networking platforms create "epistemic destabilization," where audiences become uncertain about what information can be trusted (Oliullah et al., 2026). Experimental studies by Ahmed et al. (2025) found that exposure to deepfake content significantly reduced trust in institutions and increased skepticism toward official communication channels, particularly where audiences lacked strong digital verification skills. Similarly, survey-based research conducted in the United Kingdom revealed that over 90% of respondents expressed concern that deepfakes would increase distrust in online information and manipulate public opinion, while many participants admitted limited confidence in their ability to detect manipulated media. Lundberg (2024) also observed that repeated circulation of deepfakes across digital

ecosystems contributes to declining confidence in media institutions because audiences increasingly perceive online communication as vulnerable to manipulation. These findings indicate that the impact of deepfakes extends beyond isolated deception incidents to broader crises of informational trust and authenticity.

Within corporate communication contexts, empirical evidence increasingly links repeated exposure to synthetic media with reputational vulnerabilities, stakeholder disengagement, and declining organizational credibility. Research on AI-driven misinformation demonstrates that stakeholders who frequently encounter manipulated executive messages, fraudulent promotional videos, or fabricated organizational statements become more cautious and skeptical toward corporate communication channels (Kietzmann et al., 2020). Studies further show that repeated deepfake exposure weakens stakeholder confidence in visual and audio communication because authentic corporate messages can subsequently be dismissed as potentially fabricated, a phenomenon described as the "liar's dividend" (Chesney & Citron, 2019). Recent studies on digital trust similarly report that organizations operating in highly networked environments face increasing communication risks because synthetic media spreads rapidly before verification mechanisms can respond effectively. However, much of the existing empirical literature has focused on political misinformation, journalism, and cybersecurity, with limited attention to how repeated deepfake exposure affects stakeholder trust within African corporate settings. Existing African studies remain largely conceptual or policy-oriented, leaving insufficient empirical evidence regarding how Kenyan stakeholders experience, interpret, and respond to frequent exposure to deepfake-driven corporate misinformation. This contextual gap justifies the need for further qualitative investigation into how recurring encounters with AI deepfakes shape stakeholder trust erosion within Kenyan corporate communication environments**.**

Sophistication of AI Deepfakes and Stakeholder Trust

Empirical research increasingly demonstrates that the sophistication and realism of AI-generated deepfakes significantly influence stakeholder perceptions of authenticity, credibility, and trust in digital communication. Advances in deep learning, particularly generative adversarial networks (GANs), have enabled the production of highly convincing synthetic audio, video, and image content that closely mimics real individuals in speech, appearance, and behavior (Goodfellow et al., 2014). Studies indicate that sophisticated deepfakes are more persuasive because they exploit visual and auditory cues commonly associated with credibility and legitimacy. According to Hwang (2020), the realism of deepfakes increases their persuasive capacity by reducing audiences' ability to identify manipulation, thereby amplifying the spread of misinformation and deception. Similarly, Vaccari and Chadwick (2020) found that individuals exposed to realistic synthetic media were more likely to perceive manipulated content as authentic, particularly when the content aligned with existing beliefs or contextual expectations. Empirical findings further suggest that high-quality deepfakes undermine confidence not only in fabricated content but also in legitimate digital communication, thereby contributing to broader crises of informational trust and authenticity across online environments.

Scholars have also examined how sophisticated deepfakes influence organizational reputation and stakeholder behavior within corporate communication contexts. Research by Chesney and Citron (2019) found that realistic executive impersonations and manipulated corporate announcements can generate immediate reputational harm before verification mechanisms intervene effectively. Similarly, Kietzmann et al. (2020) argue that the increasing sophistication of AI-generated media presents major risks to organizations because stakeholders often rely on visual realism, vocal accuracy, and contextual coherence as indicators of authenticity. Experimental studies conducted by Appel et al. (2022) revealed that participants exposed to highly realistic deepfake videos expressed greater uncertainty regarding the credibility of digital communication and demonstrated lower trust in institutional messaging overall. Other empirical studies indicate that sophisticated deepfakes are especially dangerous in high-information digital ecosystems because rapid dissemination through social media platforms allows manipulated content to shape stakeholder perceptions before fact-checking occurs (Westerlund, 2019). This creates significant challenges for organizations attempting to maintain credibility, manage crises, and protect stakeholder confidence within increasingly mediated communication environments.

Despite growing scholarly interest in deepfake sophistication, existing empirical studies remain concentrated in Western political, media, and cybersecurity contexts, with limited focus on corporate communication in

developing economies. In African contexts, research examining the organizational implications of sophisticated synthetic media remains relatively underdeveloped. Emerging studies in Kenya suggest that communication practitioners are increasingly concerned about the realism and rapid dissemination of AI-generated content, particularly its ability to damage reputations and undermine public trust before organizations can respond effectively (Karanja & Nzau, 2026). However, much of the available literature remains conceptual, descriptive, or policy-oriented, with limited empirical evidence regarding how stakeholders interpret sophisticated deepfakes within everyday corporate interactions. Furthermore, few studies have examined how contextual factors such as digital literacy, mobile communication dependence, and weak verification systems shape stakeholder vulnerability to highly realistic synthetic media in African corporate environments. Consequently, there remains a significant empirical gap regarding how the sophistication of AI deepfakes influences stakeholder trust erosion within Kenyan corporate communication settings. This study therefore seeks to address this gap by exploring how stakeholders perceive and respond to increasingly sophisticated forms of synthetic media in organizational communication processes**.**

Detection Difficulty of AI Deepfakes and Stakeholder Trust Erosion

Empirical studies increasingly identify detection difficulty as one of the most significant factors contributing to stakeholder distrust in digital communication environments. As AI-generated deepfakes become more sophisticated, stakeholders experience growing challenges distinguishing authentic communication from manipulated content. Research by David Green and John Swets' Signal Detection framework has frequently been applied in misinformation studies to explain how uncertainty and perceptual limitations affect decision-making under conditions of informational ambiguity. Recent empirical evidence demonstrates that most individuals struggle to identify high-quality deepfakes accurately, particularly when synthetic media contains realistic visual, vocal, and contextual cues. A large-scale experimental study by Groh et al. (2022) found that participants correctly identified deepfake videos only slightly better than chance, highlighting substantial human vulnerability to AI-generated deception. Similarly, research by Vaccari and Chadwick (2020) revealed that audiences exposed to manipulated media frequently expressed uncertainty regarding the authenticity of online information, contributing to declining confidence in digital communication systems. These findings suggest that detection difficulty not only facilitates deception but also generates broader skepticism toward legitimate organizational communication and digital evidence more generally.

Scholars further argue that technological limitations, cognitive overload, and low digital literacy intensify the difficulty of detecting deepfakes in rapidly evolving communication environments. Empirical studies indicate that stakeholders often rely on heuristic processing and superficial cues when evaluating online content, especially within high-volume social media ecosystems characterized by speed and information saturation (Pennycook & Rand, 2021). Consequently, sophisticated deepfakes can bypass critical scrutiny before verification occurs. Research by Westerlund (2019) found that the inability to verify manipulated content contributes to what he describes as "communicative uncertainty," where audiences begin questioning the credibility of both authentic and fabricated information. Similarly, studies on misinformation detection demonstrate that repeated exposure to undetected synthetic media weakens trust in institutions, media organizations, and corporate communication channels because stakeholders perceive digital environments as increasingly unreliable (Diakopoulos & Johnson, 2021). Experimental research conducted by Ahmed et al. (2025) further established that limited awareness of AI-generated manipulation significantly increased participant susceptibility to deceptive content and reduced confidence in official communication sources. These findings indicate that detection difficulty has both cognitive and organizational implications, particularly in contexts where verification systems remain underdeveloped.

Within African and Kenyan contexts, empirical research examining deepfake detection remains limited despite growing concerns regarding AI-enabled misinformation and cyber fraud. Emerging evidence suggests that many organizations across developing economies lack adequate technological infrastructure, AI literacy, and institutional preparedness to identify or respond effectively to sophisticated synthetic media threats (African Union, 2023). In Kenya, increasing dependence on digital communication platforms for organizational engagement has heightened vulnerability to manipulated content circulating through WhatsApp, Facebook, and Telegram. Karanja and Nzau (2026) found that communication practitioners in Kenya identified limited

technical expertise and inadequate detection tools as major barriers to managing AI deepfake crises effectively. However, existing studies remain largely exploratory and concentrate primarily on media practice or policy concerns rather than stakeholder trust within corporate communication settings. Furthermore, little empirical research has examined how stakeholders interpret unverifiable organizational content or how detection difficulties shape reputational perceptions and communication credibility in Kenyan corporations. This contextual gap underscores the need for qualitative investigation into how technological and human limitations associated with deepfake detection contribute to stakeholder trust erosion within Kenya's evolving digital communication landscape**.**

Ethical Dilemmas and Stakeholder Trust in Corporate Communication

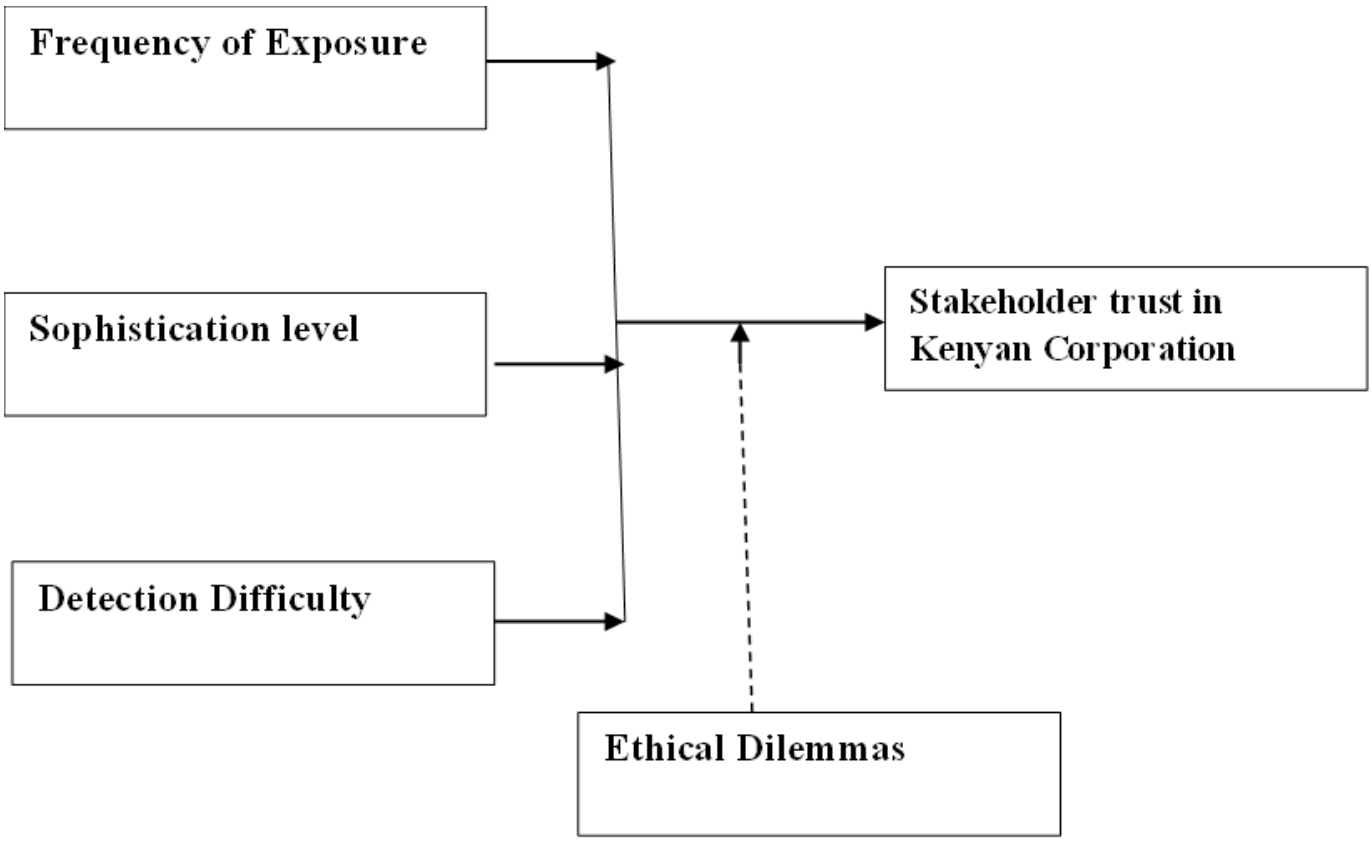
Empirical scholarship on AI ethics and corporate communication increasingly demonstrates that ethical dilemmas significantly influence stakeholder trust, organizational legitimacy, and perceptions of communication credibility. Ethical dilemmas emerge when organizations must balance technological innovation, operational efficiency, and strategic communication objectives against responsibilities related to transparency, accountability, privacy, and informational authenticity. Research by Floridi et al. (2018) argues that AI technologies generate ethical complexities because organizations often adopt advanced digital systems faster than governance structures can regulate them effectively. Within communication environments, these dilemmas become more pronounced when synthetic media technologies such as deepfakes are used without clear disclosure, verification, or accountability mechanisms. Empirical studies indicate that stakeholders are more likely to distrust organizations perceived as lacking ethical transparency or responsible AI governance practices (Martin, 2019). Similarly, corporate communication research consistently demonstrates that ethical conduct and transparent information disclosure are central determinants of stakeholder confidence, reputational stability, and organizational legitimacy (Cornelissen, 2020). Consequently, ethical dilemmas associated with AI-generated communication increasingly represent strategic risks for organizations operating within digitally mediated environments.

Existing empirical evidence further suggests that organizational responses to AI-related ethical concerns directly influence stakeholder perceptions and crisis management outcomes. Studies examining digital misinformation and AI governance reveal that organizations with proactive ethical frameworks, clear disclosure policies, and transparent communication strategies are better positioned to maintain stakeholder trust during reputational crises (Diakopoulos, 2019). Conversely, organizations perceived as evasive, unprepared, or ethically negligent experience intensified skepticism and reputational backlash following misinformation incidents. Research by Gillespie (2020) found that accountability gaps in digital communication systems contribute to declining public trust because stakeholders increasingly expect organizations to verify, monitor, and responsibly manage online information. Similarly, studies on AI ethics indicate that inadequate governance structures create uncertainty regarding consent, data privacy, and authenticity, thereby weakening stakeholder confidence in organizational decision-making processes (Taddeo & Floridi, 2018). Empirical investigations also show that ethical leadership, institutional transparency, and stakeholder-centered communication practices play important roles in mitigating distrust associated with emerging AI technologies. These findings suggest that ethical dilemmas are not peripheral concerns but central determinants of trust formation and organizational credibility in contemporary corporate communication.

Despite increasing international scholarship on AI ethics, limited empirical research has examined how ethical dilemmas shape stakeholder trust within African corporate communication contexts. Existing African studies primarily emphasize regulatory frameworks, cybersecurity concerns, or broader digital governance debates while giving insufficient attention to organizational communication ethics and stakeholder perceptions. In Kenya, emerging research indicates that organizations face growing ethical challenges related to misinformation management, AI governance, and digital transparency within rapidly expanding online communication environments (Karanja & Nzau, 2026). However, empirical evidence examining how stakeholders interpret organizational ethical responses to deepfake incidents remains scarce. Moreover, few studies have explored how ethical dilemmas such as disclosure obligations, accountability failures, transparency gaps, and verification responsibilities moderate the relationship between AI deepfakes and stakeholder trust erosion in corporate settings. Most available studies remain conceptual, policy-oriented, or focused on political misinformation rather

than routine organizational communication. This creates a significant empirical and contextual gap regarding the role of ethical dilemmas in shaping stakeholder trust within Kenyan corporations. The present study therefore seeks to address this gap by examining how ethical challenges associated with AI deepfakes influence organizational credibility and stakeholder trust in Kenya's evolving digital communication environment.

CONCEPTUAL FRAMEWORK



METHODOLOGY

This study adopted a qualitative research design to explore the influence of AI deepfakes on stakeholder trust in corporate communication within Kenyan organizations. A qualitative approach was deemed appropriate because the study sought to obtain an in-depth understanding of participants' experiences, perceptions, and interpretations regarding the ethical and communicative implications of synthetic media in organizational settings. Unlike quantitative approaches that focus on measurement and statistical relationships, qualitative research facilitates exploration of complex social phenomena within their natural contexts. The study was guided by an interpretivist paradigm, which assumes that reality is socially constructed and that individuals develop meanings based on their interactions and experiences. Given the emerging nature of AI deepfakes and the subjective dimensions of stakeholder trust, authenticity, and ethical decision-making, a qualitative approach provided the flexibility necessary to capture rich and nuanced insights into how stakeholders perceive and respond to deepfake-related communication challenges in Kenya's evolving digital environment.

The study targeted stakeholders drawn from selected Kenyan corporations operating in communication-intensive sectors, including banking, telecommunications, insurance, retail, and technology. These sectors were selected because they rely heavily on digital communication channels for customer engagement, investor relations, crisis communication, internal coordination, and corporate branding. A total of 25 participants were purposively selected, comprising communication professionals, public relations practitioners, managerial personnel, employees, and customers with relevant experience in organizational communication and digital media engagement. Purposive sampling was employed to identify information-rich participants capable of providing meaningful insights into the influence of AI deepfakes on stakeholder trust. Participants were selected based on their involvement in communication-related activities, familiarity with digital platforms, and exposure to issues

surrounding misinformation, social media communication, or AI-related technologies. The sample size was considered adequate because qualitative research prioritizes depth and richness of information rather than statistical representativeness. According to John W. Creswell and Poth (2018), qualitative studies commonly involve between 20 and 30 participants when the objective is to generate detailed insights into participants' lived experiences. Similarly, Guest, Bunce, and Johnson (2006) argue that thematic saturation is often achieved within relatively small but information-rich samples, making the selected sample appropriate for the objectives of the study.

Data were collected through semi-structured interviews, focus group discussions, and non-participant observations. Semi-structured interviews were chosen because they enabled participants to discuss their experiences and perceptions freely while allowing the researcher to probe for clarification and deeper understanding. Focus group discussions facilitated interaction among participants and generated diverse perspectives regarding the influence of AI deepfakes on organizational communication and stakeholder trust. Non-participant observations complemented interview and focus group data by providing additional contextual insights into communication practices, stakeholder interactions, and organizational responses to digital communication challenges. The use of multiple data collection methods enabled the researcher to capture different dimensions of the phenomenon under investigation and obtain a more comprehensive understanding of stakeholder experiences. Data collection was conducted over a period sufficient to ensure adequate coverage of the study objectives and continued until no substantially new themes emerged from participant responses.

To enhance the trustworthiness and credibility of the findings, several quality assurance measures were implemented throughout the research process. Methodological triangulation was achieved through the use of interviews, focus groups, and observations, allowing findings from one source to be corroborated by evidence from other sources. Triangulation strengthened the credibility of the study by minimizing the limitations associated with relying on a single method of data collection. In addition, member checking was conducted by sharing preliminary interpretations and thematic summaries with selected participants to confirm the accuracy of the researcher's interpretations. Detailed field notes were maintained during data collection to document contextual observations, participant interactions, and emerging analytical insights. These procedures enhanced the credibility, dependability, confirmability, and transferability of the findings by ensuring transparency and consistency throughout the research process.

The collected data were analyzed using thematic analysis following the framework proposed by Braun and Clarke (2006). The analysis began with familiarization, whereby interview transcripts, focus group records, and observation notes were reviewed repeatedly to gain a comprehensive understanding of the data. This was followed by the generation of initial codes that captured significant ideas, concepts, and patterns relevant to the study objectives. Similar codes were subsequently grouped into broader categories and themes, which were reviewed and refined through an iterative analytical process. The analysis resulted in the identification of four major themes: frequency of exposure to AI deepfakes, sophistication of deepfakes, detection difficulty, and ethical dilemmas in corporate communication. Thematic analysis was selected because of its flexibility and effectiveness in identifying recurring patterns while preserving the contextual richness of participants' experiences. The resulting themes provided the basis for interpreting how AI deepfakes influence stakeholder trust within Kenyan corporate communication contexts.

Ethical considerations were observed throughout the study to protect participants and maintain the integrity of the research process. Participation was voluntary, and informed consent was obtained from all respondents before data collection commenced. Participants were informed about the purpose of the study, the nature of their involvement, and their right to withdraw from the study at any stage without penalty. Confidentiality and anonymity were guaranteed by ensuring that personal identifiers were excluded from transcripts, field notes, and the final report. Data were stored securely and used exclusively for academic purposes. Furthermore, findings are presented in a manner that protects participant identities while accurately reflecting their views and experiences. By adhering to established ethical principles and rigorous qualitative research procedures, the study generated credible insights into the influence of AI deepfakes on stakeholder trust and the moderating role of ethical dilemmas within Kenya's corporate communication landscape.

ANALYSIS AND RESULTS

Frequency of Exposure

The findings of the study revealed that repeated exposure to AI-generated deepfakes significantly contributes to the erosion of stakeholder trust in Kenyan corporate communication. Participants across interviews and focus group discussions consistently indicated that the increasing circulation of manipulated videos, audio clips, and fabricated digital messages on social media platforms has created a climate of uncertainty regarding the authenticity of organizational communication. Most respondents identified platforms such as WhatsApp, Facebook, TikTok, Telegram, and X as major channels through which synthetic media spreads rapidly before verification can occur. Several participants observed that the high frequency of exposure to misleading content has gradually normalized suspicion toward digital corporate messages. One communication manager from a Nairobi-based financial institution stated, "Today people first doubt before they believe. Even genuine communication from companies is questioned because fake videos and voice recordings have become common online." This finding suggests that repeated encounters with synthetic media have altered how stakeholders interpret organizational communication, resulting in heightened skepticism and declining confidence in official corporate messaging.

The study further established that frequent exposure to deepfakes affects stakeholder behavior and organizational relationships in significant ways. Participants reported that consumers increasingly delay decision-making until information is independently verified, while employees and investors often express caution toward announcements circulated through digital platforms. Several respondents explained that even after clarification is issued, reputational damage frequently persists because manipulated content spreads faster than corrections. A public relations practitioner observed that "by the time an organization explains that something is fake, the damage has already happened because people have already shared the clip thousands of times." These findings indicate that repeated exposure not only weakens confidence in specific communication but also contributes to broader distrust toward institutional communication systems. The findings correspond with the arguments of Vaccari and Chadwick (2020), who found that sustained exposure to manipulated digital content cultivates long-term distrust in online communication. Similarly, Westerlund (2019) argues that deepfakes generate a "liar's dividend" in which audiences become uncertain about the credibility of both authentic and fabricated information. The current study extends these findings to the Kenyan corporate context, where high mobile penetration and widespread social media usage intensify the speed and frequency of exposure to synthetic media.

Another important finding was that repeated exposure to deepfakes contributes to emotional fatigue and communication disengagement among stakeholders. Participants explained that constant encounters with manipulated content create frustration, anxiety, and confusion, especially where verification mechanisms are weak or inconsistent. Some respondents admitted that they had begun ignoring online corporate communication entirely unless confirmed through multiple channels. This growing distrust was particularly evident among younger stakeholders who consume large volumes of digital content daily. One participant noted, "You cannot trust everything you see anymore. Sometimes it is easier to ignore company updates completely than risk believing fake information." Such responses indicate that the consequences of repeated deepfake exposure extend beyond isolated misinformation incidents to broader disengagement from organizational communication processes. These findings support cultivation-based perspectives suggesting that continuous exposure to deceptive media shapes long-term perceptions and behavioral attitudes toward communication systems (Morgan et al., 2015). However, unlike most existing studies that focus primarily on political misinformation and journalism, this study demonstrates that the cumulative effects of synthetic media exposure also pose substantial risks to stakeholder trust, organizational reputation, and relationship management within Kenyan corporate environments.

Sophistication Level

The study established that the sophistication and realism of AI-generated deepfakes significantly intensify stakeholder distrust in corporate communication. Participants consistently emphasized that advances in synthetic media technology have made manipulated content increasingly difficult to distinguish from authentic

communication. Respondents identified visual realism, voice imitation, facial synchronization, and contextual accuracy as major factors contributing to the persuasive power of deepfakes. Several participants explained that modern AI-generated content often incorporates local accents, familiar corporate branding, and culturally recognizable communication styles, thereby increasing perceived authenticity among audiences. One participant working in corporate communications stated, "Some deepfakes now sound exactly like company executives. They even use familiar phrases and speaking styles, making it very hard for ordinary people to notice manipulation." These findings indicate that the realism of deepfakes enhances their credibility and allows deceptive content to spread rapidly before organizations can intervene effectively.

The findings further revealed that sophisticated deepfakes create severe reputational and financial risks for organizations. Participants described instances in which manipulated videos or fabricated executive statements triggered panic among customers, uncertainty among investors, and confusion among employees. Respondents noted that stakeholders frequently rely on visual and auditory cues when evaluating organizational credibility, making them particularly vulnerable to realistic synthetic media. One banking sector participant explained that "when people hear a familiar executive voice or see a realistic video, they immediately assume it is genuine, especially if it appears on social media." Such findings demonstrate that highly sophisticated deepfakes exploit existing trust structures within corporate communication systems. The findings align with Chesney and Citron (2019), who argue that realistic synthetic media undermines confidence in digital evidence by manipulating the same cues audiences traditionally associate with authenticity. Similarly, Hwang (2020) found that highly convincing deepfakes increase susceptibility to misinformation because audiences tend to process realistic content through peripheral rather than analytical evaluation. The current study confirms these concerns within Kenyan corporate settings, where stakeholders increasingly depend on digital communication for organizational updates, investment decisions, and customer engagement.

Participants also highlighted that the sophistication of deepfakes complicates organizational crisis management efforts. Several respondents explained that organizations often struggle to respond quickly because highly realistic synthetic media initially appears credible even to internal communication teams. This delay in identifying manipulation allows misinformation to spread extensively across digital networks before corrective action is taken. Participants further indicated that sophisticated deepfakes create long-term reputational uncertainty because stakeholders continue questioning whether future communication may also be fabricated. One respondent observed that "even after clarifications are made, people still remain suspicious because the fake content looked very real." These findings suggest that sophisticated deepfakes do not merely create temporary misinformation crises but contribute to sustained distrust toward corporate communication systems. The findings support the Elaboration Likelihood Model advanced by Petty and Cacioppo (1986), which suggests that audiences frequently rely on superficial indicators such as realism, familiarity, and perceived credibility when processing information in digitally saturated environments. However, unlike much of the existing literature concentrated in Western political and media contexts, this study demonstrates that the sophistication of deepfakes also presents significant organizational communication challenges in African corporate environments characterized by rapid digital adoption and uneven verification capabilities.

Detection Difficulty

The study identified detection difficulty as one of the most significant factors contributing to stakeholder trust erosion in Kenyan corporate communication. Participants consistently reported that distinguishing authentic communication from manipulated synthetic media has become increasingly difficult due to rapid technological advancements in AI-generated content. Respondents emphasized that many organizations lack adequate tools, expertise, and institutional preparedness to identify sophisticated deepfakes effectively. Communication professionals noted that even trained personnel sometimes struggle to verify whether digital content has been manipulated, especially when deepfakes involve realistic facial expressions, synchronized speech, and professionally edited visuals. One participant stated, "Sometimes the fake content is so convincing that even communication teams debate whether it is genuine before confirming." These findings demonstrate that detection challenges extend beyond ordinary stakeholders and affect organizations themselves, thereby increasing communication uncertainty and reputational vulnerability.

Participants further explained that human limitations significantly intensify the difficulty of detecting deepfakes. Several respondents referred to cognitive overload caused by the large volume of information circulating on social media platforms daily. Because stakeholders are constantly exposed to online content, many rely on speed and familiarity rather than careful verification when interpreting information. Respondents observed that emotionally charged or sensational content spreads especially quickly because audiences often react impulsively before evaluating credibility. One participant from a telecommunications company noted, "People rarely stop to verify information before sharing it. If something looks believable and emotional, it spreads immediately." These findings align with Pennycook and Rand (2021), who found that individuals frequently depend on heuristic thinking rather than analytical reasoning when consuming digital information. Similarly, Groh et al. (2022) established that most people perform poorly when attempting to distinguish authentic videos from deepfakes, particularly when manipulated content contains realistic contextual cues. The current findings therefore suggest that detection difficulty is both a technological and psychological challenge affecting stakeholder trust within digital corporate communication systems.

The findings also revealed that weak verification mechanisms and limited AI literacy exacerbate communication breakdowns in Kenyan organizations. Participants observed that many corporations lack standardized protocols for authenticating digital communication during crises involving manipulated media. In some organizations, uncertainty regarding who should verify information or respond publicly contributed to delayed reactions and stakeholder confusion. Respondents further explained that limited public awareness regarding deepfake technologies increases vulnerability to deception because many stakeholders remain unfamiliar with how synthetic media is created or detected. One respondent remarked, "Most people know fake news exists, but they do not understand how advanced AI-generated videos have become." These findings correspond with Signal Detection Theory, which explains that uncertainty increases when individuals lack sufficient sensitivity or criteria to distinguish signals from noise (Green & Swets, 1966). The findings also support recent studies by Vaccari and Chadwick (2020), which suggest that detection difficulties contribute to generalized distrust toward online communication systems. However, unlike existing studies focusing largely on political misinformation, this study demonstrates that verification challenges significantly affect organizational legitimacy, stakeholder engagement, and corporate communication credibility within Kenya's evolving digital business environment.

Ethical Dilemmas as Moderator

The findings of the study revealed that ethical dilemmas play a significant moderating role in shaping how AI deepfakes influence stakeholder trust in Kenyan corporate communication. Participants consistently identified transparency gaps, accountability challenges, disclosure failures, and regulatory uncertainty as major ethical concerns complicating organizational responses to synthetic media threats. Many respondents argued that stakeholder reactions to deepfake incidents often depend not only on the manipulated content itself but also on how organizations respond during communication crises. Participants indicated that organizations perceived as evasive, secretive, or unprepared tend to experience greater reputational damage and stakeholder skepticism. One participant observed, "People lose trust faster when a company appears to hide information or delay communication after a fake video circulates." These findings suggest that ethical conduct and communication transparency significantly influence whether organizations mitigate or intensify distrust following deepfake-related incidents.

The study further established that proactive ethical leadership and responsible AI governance can reduce the negative effects of deepfakes on stakeholder trust. Participants from organizations with established communication protocols reported greater confidence in crisis response processes and stronger stakeholder reassurance during misinformation incidents. Several respondents emphasized the importance of timely clarification, transparent disclosure, and internal verification procedures in restoring credibility after manipulated content emerges. One corporate affairs officer explained, "Our organization now verifies sensitive communication through multiple channels and responds quickly whenever suspicious content appears online." Participants also highlighted the importance of employee awareness training and digital literacy initiatives in strengthening institutional resilience against synthetic media threats. These findings support Floridi et al. (2018), who argue that ethical AI governance frameworks are necessary for maintaining accountability and public trust in digital systems. Similarly, Martin (2019) found that organizations demonstrating ethical responsibility and

transparent communication practices are more likely to sustain stakeholder confidence during technological crises. The current study therefore indicates that ethical preparedness can moderate the relationship between deepfakes and trust erosion by improving organizational responsiveness and communication credibility.

However, participants also reported that many Kenyan organizations still lack comprehensive ethical guidelines regarding AI usage and digital communication management. Several respondents expressed concern that regulatory frameworks remain underdeveloped, leaving organizations uncertain about legal responsibilities and accountability standards in relation to deepfake incidents. Others argued that ethical considerations are often treated reactively rather than proactively, with organizations responding only after reputational harm has occurred. One participant stated, "Most companies are still learning about AI risks after incidents happen instead of preparing early." These findings reveal significant institutional and governance gaps within Kenya's corporate communication environment. They further suggest that ethical dilemmas extend beyond technical concerns to broader questions regarding organizational legitimacy, stakeholder protection, and responsible digital governance. While previous studies have primarily focused on legal and policy dimensions of AI ethics, the present findings demonstrate that ethical dilemmas also directly shape stakeholder perceptions, communication trust, and crisis response effectiveness in organizational contexts. Consequently, the study highlights the importance of integrating ethical governance, transparent communication strategies, and stakeholder-centered AI policies into corporate communication frameworks to strengthen organizational resilience against emerging synthetic media threats.

CONCLUSION

This study demonstrates that AI-generated deepfakes pose a significant threat to stakeholder trust within Kenyan corporate communication environments. The findings reveal that repeated exposure to synthetic media, increasing sophistication of deepfake technologies, and persistent detection difficulties collectively contribute to heightened skepticism, reputational instability, and declining confidence in organizational communication. In digitally connected environments where stakeholders rely heavily on online platforms for information, manipulated content undermines perceptions of authenticity and weakens institutional credibility. The study further establishes that ethical dilemmas—including transparency gaps, weak accountability structures, delayed disclosure, and limited regulatory preparedness—intensify the negative effects of deepfakes on stakeholder trust. However, the findings also indicate that proactive ethical leadership, transparent communication practices, employee awareness initiatives, and effective verification mechanisms can mitigate these risks and strengthen organizational resilience. By providing empirical evidence from Kenyan corporate settings, the study addresses existing contextual gaps in African scholarship on AI deepfakes and corporate communication. The research therefore underscores the need for stakeholder-centered AI governance frameworks that prioritize authenticity, accountability, and responsible digital communication in increasingly complex information ecosystems.

RECOMMENDATIONS

Kenyan corporations should strengthen organizational preparedness against AI deepfakes by investing in continuous AI ethics and digital verification training for communication teams, executives, and employees. Regular training programs can improve awareness of synthetic media threats, enhance detection capabilities, and reduce vulnerability to manipulated corporate content. Organizations should also promote ethical leadership practices centered on transparency, accountability, and responsible AI governance to reinforce stakeholder confidence during digital communication crises. Clear disclosure policies regarding AI-generated content and timely stakeholder engagement are essential for maintaining organizational credibility and trust.

Additionally, corporations should establish standardized verification frameworks incorporating watermarking technologies, metadata authentication, multi-channel confirmation systems, and real-time monitoring tools to detect and respond to deepfake threats more effectively. Investment in localized AI detection technologies suited to Kenya's digital environment is necessary to address technological limitations and improve accessibility across organizations. Firms should further collaborate with regulators, technology providers, cybersecurity experts, and industry associations to develop coordinated responses, information-sharing mechanisms, and crisis simulation exercises. Public digital literacy initiatives targeting employees, consumers, and stakeholders are equally

important in improving awareness of synthetic media manipulation. Collectively, these strategies will strengthen organizational resilience, enhance communication authenticity, reduce stakeholder distrust, and support responsible adoption of AI technologies within Kenya's evolving corporate communication landscape.

REFERENCES

1. African Union. (2023). *African Union digital transformation and cybersecurity framework*. African Union Commission.
2. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211
3. Amatika, F. (2022). The regulation of deepfakes in Kenya. *Journal of Intellectual Property and Information Technology Law, 2*(1), 145–186.
4. Appel, G., Neelbauer, J., & Schweidel, D. A. (2022). Generative AI has an intellectual property problem. *Harvard Business Review, 100*(5), 78–85.
5. Business Daily Africa. (2026, February 2). Implications of deepfake surge for the legal and business landscape. Business Daily Africa
6. Chesney, R., & Citron, D. K. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs, 98*(1), 147–155.
7. CIO Africa. (2026, February 5). Deepfake scams, AI malware raise cyber risks in Kenya. CIO Africa
8. Communications Authority of Kenya. (2024). *First quarter sector statistics report for the financial year 2023/2024*. Communications Authority of Kenya.
9. Cornelissen, J. (2020). *Corporate communication: A guide to theory and practice* (6th ed.). Sage Publications.
10. Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
11. Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in communication. *Digital Journalism, 9*(8), 1051–1062. https://doi.org/10.1080/21670811.2021.1903978
12. ESET. (2025). *Deepfake scams and AI malware raise cyber risks in Kenya*. ESET Threat Report.
13. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
14. Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.
15. Gerbner, G., & Gross, L. (1976). Living with television: The violence profile. *Journal of Communication, 26*(2), 172–199. https://doi.org/10.1111/j.1460-2466.1976.tb01397.x
16. Gillespie, T. (2020). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*, 2672–2680.
18. Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
19. Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Human detection of machine manipulated media. *Proceedings of the National Academy of Sciences, 119*(13), e2110013119. https://doi.org/10.1073/pnas.2110013119
20. GSMA. (2023). *The mobile economy Sub-Saharan Africa 2023*. GSM Association.
21. Hussein, K., & Özad, B. E. (2025). AI-driven media manipulation: Public awareness, trust, and digital media in the age of deepfakes. *International Journal of Media and Communication Studies*.
22. Hwang, Y. (2020). The rise of deepfake technology and its implications for media and communication. *Journal of Digital Media Ethics, 5*(2), 45–59.
23. Iyer, C. A. V., & Gokhale, N. P. (2026). Artificial intelligence, deepfakes, and corporate intelligence: A systematic literature review on ethical and strategic implications for business transparency and decision-making. In *Artificial intelligence and business intelligence in corporate ecosystems* (pp. 71–95). Emerald Publishing. https://doi.org/10.1108/978-1-80592-911-620261004

24. Karanja, J. M. (2023). *Africa deepfakes buzz: Exploring online media socio-political discourses on deepfakes – Focus on Kenya, Nigeria and South Africa*. OSF.
25. Karanja, J. M., & Nzau, T. (2026). The emergence of AI deepfakes and their impact on strategic crisis communication: Practitioners' perspectives from Kenya. *Journal of Communication Management*. https://doi.org/10.1111/1468-5973.70140
26. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006
27. Kigwiru, V. K. (2022). Deepfake technology and elections in Kenya: Can legislation combat the harm posed by deepfakes? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4229272
28. Kitchen, P. J., Kerr, G., Schultz, D. E., McColl, R., & Pals, H. (2014). The elaboration likelihood model: Review, critique and research agenda. *European Journal of Marketing, 48*(11/12), 2033–2050. https://doi.org/10.1108/EJM-12-2011-0776
29. Lundberg, A. (2024). Deepfakes and declining trust in digital communication ecosystems. *AI & Society, 39*(2), 455–468. https://doi.org/10.1007/s00146-024-02072-1
30. Mare, A. (2020). Social media, politics and communication in Africa. *African Journalism Studies, 41*(1), 1–7. https://doi.org/10.1080/23743670.2020.1729838
31. Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3
32. Morgan, M., Shanahan, J., & Signorielli, N. (2015). Yesterday's new cultivation, tomorrow. *Mass Communication and Society, 18*(5), 674–699. https://doi.org/10.1080/15205436.2015.1072725
33. Mutsvairo, B., & Ragnedda, M. (2019). Mapping the digital divide in Africa. In M. Ragnedda & B. Mutsvairo (Eds.), *Mapping the digital divide in Africa* (pp. 1–12). Amsterdam University Press.
34. Oliullah, M., Rashid, M., & Khan, S. (2026). Epistemic destabilization and synthetic media exposure in digital ecosystems. *Computers in Human Behavior Reports, 14*, 100402.
35. Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences, 25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007
36. Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.
37. Smile ID. (2025). *Digital identity fraud in Africa report 2025*. Smile Identity.
38. Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. *Nature, 556*(7701), 296–298. https://doi.org/10.1038/d41586-018-04602-6
39. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust. *Social Media + Society, 6*(1), 1–13. https://doi.org/10.1177/2056305120903408
40. Vecchietti, G. (2025). Managing deepfakes with artificial intelligence: Introducing a business privacy calculus model. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2024.105149
41. Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 9*(11), 39–52. https://doi.org/10.22215/timreview/1282
42. World Economic Forum. (2024). *Global risks report 2024*. World Economic Forum
43. Zerfass, A., Verčič, D., Nothhaft, H., & Werder, K. P. (2022). Strategic communication and organizational trust in digital environments. *International Journal of Strategic Communication, 16*(3), 145–161. https://doi.org/10.1080/1553118X.2022.2048691