CURRENT RESEARCH IN
COMMUNICATION

VOLUME 1-1| 2026 | PAGES 55-67 | DOI: 10.66699/4arhqa29

RESEARCH ARTICLE

Beyond Denial: Toward a Relational Verification Framework for Public Relations Responses to Deepfake Crises in Kenya

Check for updates

Mercy Ndirangu*[1], Millicent Oyier[2], Olive Metet[3], Marcella Onsumu[4]

[1*] Author for Correspondence, PhD candidate in Communication Studies, Daystar University; intergenerational and crisis communication.
[2 3] PhD candidates in Communication Studies, Daystar University; intergenerational and crisis communication.

ABSTRACT

Artificial Intelligence has introduced a new problem into crisis communication: synthetic audio-visual content that audiences cannot distinguish from authentic record. Kenya, combining high mobile penetration, low institutional trust and a contested electoral cycle, is an early site of this disruption. This qualitative documentary case study analyzes six deepfake incidents verified by Africa Check and PesaCheck in 2025, triangulated with media and official sources, and coded against Situational Crisis Communication Theory (SCCT), Framing Theory and Media Ecology Theory through reflexive thematic analysis. Kenyan deepfakes proved provenance-parasitic, grafting synthetic audio onto authentic footage and borrowing institutional credibility markers. Fact-checkers performed almost all correction; no targeted principal issued a documented rebuttal. Institutions were also vectors, as when a Principal Secretary circulated a synthetic broadcast. Corrections were channel-asymmetric, in English on websites while fabrications spread in Kiswahili and Sheng on social platforms. The study argues that SCCT's victim-cluster prescriptions are necessary but insufficient in low-trust environments, and proposes the Relational Verification Framework (anticipatory framing, contextual attribution, relational amplification and trust aftercare), anchored in Sawubona philosophy, which treats verification as a relational rather than institutional accomplishment.

Keywords: deepfakes; crisis communication; public relations; AI-generated misinformation; Kenya; SCCT; media ecology; relational verification; Sawubona

HOW TO CITE

Ndirangu, M., Oyier, M., Metet, O., & Onsumu, M. (2026). Beyond Denial: Toward a Relational Verification Framework for Public Relations Responses to Deepfake Crises in Kenya. *Current Research in Communication*, 1(1), 55-67. https://doi.org/10.66699/4arhqa29

Open access under the Creative Commons Attribution 4.0 International license: creativecommons.org/licenses/by/4.0/

Introduction

In the first week of July 2025, as Kenya absorbed the aftermath of a violent protest cycle, videos circulated on TikTok showing President William Ruto appearing to announce his resignation. In one clip, he says he is tired, has failed to rule the country, and accepts a single term; in another, he concedes that Kenyans have said he must go. Neither statement was made. The footage was taken from an authentic engagement with delegates from Kisumu County held at the President's Nairobi residence on 18 June, with synthetic audio superimposed. By the time the first clip was removed from the platform, it had been viewed more than 1.4 million times (Africa Check, 2025b).

The incident is instructive less for its technical sophistication than for what did not happen around it. No rebuttal was issued by the Office of the President or the Government Spokesperson during the window when the videos circulated most widely. Correction, when it came, came from a fact-checking organization, in English, on a website, after the fabrication had already saturated a Kiswahili-speaking TikTok audience. The crisis was not a failure of detection. It was a failure of crisis communication.

Crisis communication as a field has always presumed a shared evidentiary floor. Whether in the reactive damage control of the early twentieth century (Coombs, 2015), the stage models of the 1980s (Fink, 1986), or the contingency logic of Situational Crisis Communication Theory (Coombs, 2007), the discipline's strategies presuppose that publics can, in principle, establish what happened and then assign responsibility for it. Deepfakes attack that floor directly. When synthetic video is indistinguishable from record, the question is no longer how much responsibility a public attributes to an organization, but whether the public can determine what occurred at all (Chesney & Citron, 2019; Rini, 2020).

Kenya is an early and unusually legible site for observing this shift. It combines dense mobile connectivity and platform-native political discourse with documented deficits in institutional and media trust, a legacy of electoral information disorder stretching from 2007 to 2022, and an approaching 2027 general election that has already begun to generate synthetic content at scale (Nation Africa, 2025). Kenyan communicators are therefore not preparing for a hypothetical threat; they are already managing it, largely without a framework for doing so.

This paper asks how that management is actually being performed, and what theory would need to change to describe it adequately. It pursues four research questions. First, what are the characteristics, targets and circulation dynamics of AI-generated deepfakes in Kenya's public sphere during 2025? Second, how did institutional and non-institutional actors respond to these incidents, and through which channels? Third, how adequately do Situational Crisis Communication Theory, Framing Theory and Media Ecology Theory account for those responses in a low-trust, high-circulation environment? Fourth, what context-specific model of public relations response can be derived from the Kenyan record?

The paper proceeds in eight further sections: a review of the literature, a statement of the theoretical framework, an account of the documentary case-study methodology, the findings, a discussion of their theoretical implications, the proposed Relational Verification Framework, limitations, and conclusions.

Literature Review

From damage control to contingency

Crisis communication emerged from public relations, risk management and organizational behaviour as an essentially reactive practice. Early twentieth-century practitioners worked through controlled statements and press releases issued to a small number of gatekeeping outlets (Coombs, 2015), an approach consistent with the linear source-to-receiver model that dominated mid-century communication theory (Shannon & Weaver, 1949). Fink (1986) introduced the first influential staging of crises into prodromal, acute, chronic and resolution phases, reframing crises as anticipable rather than merely disruptive and establishing preparedness as a legitimate object of planning.

Situational Crisis Communication Theory marked the field's decisive move toward contingency. Coombs (2007) argued that response strategy should be selected according to the degree of crisis responsibility that stakeholders attribute to the organization, sorting crises into victim, accidental and preventable clusters and matching each to a repertoire of denial, diminishment, rebuilding and bolstering strategies. The theory's enduring influence rests on a specific and, for present purposes, consequential assumption: that the contested variable in a crisis is the *attribution of responsibility for an event whose occurrence is not itself in dispute*.

The digital turn and polyphonic crisis environments

Networked media dissolved the gatekeeping structure that the classical model assumed. Jin et al. (2014) demonstrated that crisis origin, information form and source jointly condition how publics respond on social platforms, while Veil et al. (2011) documented the corresponding shift toward real-time monitoring and dialogic engagement. Frandsen and Johansen (2017) characterized the resulting condition as multivocal: crises are narrated simultaneously by organizations, journalists, activists and ordinary users, none of whom controls the aggregate account. Wardle and Derakhshan (2017) supplied the now-standard vocabulary distinguishing misinformation, disinformation and malinformation by intent and veracity.

Deepfakes and the epistemic threat

Synthetic media introduces a distinct order of problem. Chesney and Citron (2019) identified not only the direct harms of fabricated audiovisual content but its second-order consequence, the liar's dividend: as public awareness of deepfakes grows, actors accused on the basis of authentic recordings acquire a credible route to dismissal. Rini (2020) formalized the underlying mechanism, arguing that recorded media has historically functioned as an epistemic backstop for testimony, and that synthetic media removes that backstop without supplying a replacement.

Empirical work suggests the dominant effect is uncertainty rather than deception. Vaccari and Chadwick (2020) found that exposure to political deepfakes did not reliably mislead participants but did significantly reduce trust in news generally. This distinction matters for practice: if the primary harm is corrosive doubt rather than specific false belief, correction alone is an insufficient remedy, since a successful debunk may deepen the very uncertainty it seeks to resolve. McIntyre (2018) situates this within a broader post-truth condition in which the authority of evidence is itself politically contested.

Kenya's information ecosystem

Kenya's exposure is structural rather than incidental. Mobile-first connectivity has produced high-velocity circulation through closed and semi-closed channels; Mudavadi et al. (2025) document how Kenyans encounter and negotiate misinformation on WhatsApp, where corrective intervention by institutions is effectively impossible and verification is performed socially, within groups, by trusted members. Mudavadi and Madrid-Morales (2024) situate Kenyan counter-disinformation practice within a wider African digital journalism context in which fact-checking capacity has grown considerably faster than institutional communication capacity.

The regulatory position is similarly uneven. The Computer Misuse and Cybercrimes Act (2018) provides a basis for action against the publication of false information, but was drafted before generative synthesis was widely available, contains no provision addressing synthetic media specifically, and sits in unresolved tension with constitutional protections for expression. At continental level, the African Union adopted its first Continental Artificial Intelligence Strategy in July 2024, covering 2025 to 2030, though implementation at member-state level remains nascent (African Union, 2024).

The research gap

Kenyan-focused scholarship on deepfakes has advanced substantially in recent years. Karanja (2026) examines the perceptions of Kenyan public relations practitioners confronting AI deepfakes, establishing

that practitioners recognize the threat while lacking protocols to address it. Wack and Prochaska (2026), working with a survey experiment of 7,000 Kenyan respondents, show that audience interpretation, and specifically the social commentary surrounding a deepfake, materially conditions its effect. Together these establish the practitioner and audience sides of the problem.

Two gaps remain. First, neither study interrogates the documented incident record itself: what Kenyan deepfakes actually did, how they were constructed, who corrected them and through which channels. The empirical archive assembled by Kenya's accredited fact-checking organizations during 2025 has not been analyzed as a corpus. Second, and more consequentially for practice, no study has yet translated these findings into an integrated response model calibrated to Kenya's trust conditions. This paper addresses both.

Theoretical Framework

The study is framed by three established theories, deliberately positioned as a set of testable expectations rather than as explanatory apparatus, together with a fourth perspective introduced to address what the first three do not reach.

Situational Crisis Communication Theory

SCCT (Coombs, 2007; Coombs & Holladay, 2012) predicts that a deepfake target falls within the victim cluster, since responsibility for the event lies wholly elsewhere, and that the appropriate response repertoire is therefore denial, clarification and attacking the accuser. The theory expects that a truthful denial from a low-responsibility actor will substantially restore reputational position. This paper treats that expectation as a proposition to be examined rather than assumed.

Framing Theory

Entman (1993) defines framing as the selection and salience of aspects of a perceived reality so as to promote a particular problem definition, causal interpretation or evaluation. Applied here, a deepfake is a framing act executed at maximum salience: it does not argue an interpretation of an event but manufactures the event itself. Framing theory predicts a first-frame advantage, whereby the initial interpretation persists unless displaced by a stronger and repeatedly asserted alternative, and supports the inoculation logic of prebunking, in which forewarning about manipulation techniques reduces susceptibility (Lewandowsky & van der Linden, 2021).

Media Ecology Theory

Postman (1970) argued that media are not neutral conduits but environments that reshape what can be communicated and how it is received. The theory directs attention away from message content toward the structural properties of the channel: its speed, its default modality, its algorithmic sorting, its language and its closure to outside intervention. It predicts that responses formulated for one environment will underperform when the crisis is occurring in another.

Sawubona and relational verification

The three theories above share a common assumption inherited from high-trust institutional contexts: that credibility is a property of sources, and that publics assess it by evaluating the source's authority, consistency and track record. In environments where institutional trust is low, this assumption does not describe practice.

The Sawubona philosophy of African communication offers a productive alternative. Rooted in the Nguni greeting rendered as *I see you*, and answered with *I am here because you see me*, Sawubona holds that personhood and, by extension, knowledge are constituted relationally rather than individually (Kamwangamalu, 1999; Mbigi, 1997). Christians (2004) develops the corresponding communitarian media

ethic, in which truth-telling is an obligation discharged within a community of relationship rather than a transmission from an authoritative centre.

Translated into the vocabulary of crisis communication, this yields a specific proposition: where trust in institutions is low, a claim's credibility derives less from *who originated it* than from *who relayed it and in what relationship that relayer stands to the recipient*. Verification, on this account, is a relational accomplishment. The paper adopts this as an analytic lens and returns to it in the model proposed in Section 7.

Methodology

Research design and paradigm

The study adopts a qualitative documentary case-study design, appropriate to an emergent phenomenon for which the analytic priority is contextual depth rather than frequency estimation (Creswell & Poth, 2018). It is situated within an interpretivist paradigm, which holds that meanings surrounding crisis events are socially constructed and must be understood within their setting (Lincoln & Guba, 1985). No human participants were involved; the design is wholly documentary.

Units of analysis and sampling

The unit of analysis is the *deepfake incident*, defined as a discrete item of synthetic or manipulated audiovisual content together with its circulation and the responses it elicited. Criterion sampling was used (Patton, 2015), applying four inclusion criteria: (i) a verdict published by Africa Check or PesaCheck, Kenya's two accredited fact-checking organizations, between 1 January and 31 December 2025; (ii) content involving synthetic or manipulated audiovisual media rather than decontextualised authentic material; (iii) a Kenyan public figure or institution as subject; and (iv) sufficient published documentation of circulation and response to permit coding across all five domains of the instrument. Six incidents satisfied all four criteria and constitute the corpus (Table 1). The corpus is not statistically representative; the claim advanced is analytic rather than statistical generalization.

Table 1

Corpus of verified Kenyan deepfake incidents, 2025

ID	Incident and date of verdict	Subject	Platform and documented reach	Verifying body
C1	Synthetic video announcing a 2027 presidential bid (4 April 2025)	Uhuru Kenyatta	TikTok; 820,000+ views, 35,000 likes, 2,300 shares	Africa Check
C2	Synthetic broadcast segment attributed to CNN on Kenya's Sudan diplomacy (21 February 2025)	Kenyan foreign policy; a foreign correspondent impersonated	X; c. 60,000 views before deletion	Kenyan news media
C3	Synthetic resignation videos following protest deaths (7 July 2025)	William Ruto	TikTok; 1.4 million+ views before removal	Africa Check
C4	Manipulated video urging opposition unity against the President	Uhuru Kenyatta	Facebook 693,000 views; TikTok 269,000 views	Africa Check
C5	Manipulated video hinting at a 2027 candidacy (3 December 2025)	Uhuru Kenyatta	Facebook 457,000 views; TikTok 1.3 million views	Africa Check
C6	Altered video describing the 2022 election as a mistake (4 December 2025)	Uhuru Kenyatta	Facebook 2.6 million views	Africa Check; PesaCheck

Note. Reach figures are as reported at the time of verification and represent minima, since content was in several cases removed or re-uploaded.

Instrumentation

A structured coding frame was developed as the study's instrument, comprising five domains derived from the theoretical framework and refined through pilot coding of two incidents falling outside the sampling window. The frame is reproduced as Table 2 to permit replication.

Table 2
Coding frame

Domain	Categories coded
D1. Artefact construction	Modality (video, audio, still); synthesis method where reported; source of underlying footage; credibility markers appropriated; detection cues cited by the verifier
D2. Circulation	Platform of origin; migration path; documented reach; language and register; elapsed time from posting to published verdict
D3. Response actor	Who responded (targeted principal, institutional communication office, fact-checker, platform, regulator, none); response latency
D4. Response strategy	Coded against SCCT clusters and strategies (denial, clarification, attacking the accuser, bolstering, corrective action, renewal); presence of supporting evidence; presence of an articulated counter-frame
D5. Channel symmetry	Correspondence between the platform, language, modality and register of the fabrication and those of the correction

Data collection procedures

Fact-check verdicts were retrieved from the published archives of Africa Check and PesaCheck and archived locally with retrieval dates. Each incident was then triangulated against at least two independent Kenyan or regional news reports and, where available, against primary institutional statements issued by the actors involved. This produced three evidentiary strata per incident: the verifier's account, contemporaneous journalistic coverage, and primary institutional communication. Divergences between strata were recorded and are reported in the findings where material.

Data analysis

Reflexive thematic analysis was applied following the six phases outlined by Braun and Clarke (2006): familiarization, initial coding, theme generation, theme review, definition, and reporting. Coding proceeded in two passes. The first was deductive, applying the categories in Table 2, derived from SCCT, Framing Theory, and Media Ecology Theory. The second was inductive, allowing patterns not anticipated by those theories to surface. The four themes reported in Section 5 emerged from the second pass and were then re-tested against the full corpus.

Trustworthiness

Credibility was pursued through the three-stratum triangulation described above, so that no finding rests on a single source. Dependability is supported by the documented coding frame, the archived corpus with retrieval dates, and a decision log recording each analytical judgement and its rationale. Transferability is addressed through thick description of Kenya's regulatory, linguistic and platform conditions, enabling readers to assess applicability to comparable settings. Confirmability rests on the public availability of every primary document, permitting independent re-coding, and on explicit reflexivity: the author is a practicing Kenyan public relations professional, a positionality that affords interpretive access to professional norms while carrying a risk of sympathetic reading of practitioner conduct, mitigated here by coding institutional non-response as a positive analytical category rather than as absent data.

Ethical considerations

The study involves no human participants and no personal data. All analyzed material was published in the public domain by fact-checking organizations and news media, and all subjects are public figures acting in

public capacities. Accordingly, institutional ethics clearance is not required for this design, and none was sought. Two ethical commitments nonetheless governed the analysis. First, no synthetic artefact is reproduced, linked, or described in sufficient detail to assist reconstruction, since scholarly circulation of such material risks extending the harm it caused. Second, individuals who were themselves targets of fabrication are named only where their identity is integral to the analysis and already a matter of extensive public record.

Findings

Coding across the five instrument domains produced a consistent descriptive picture before thematic interpretation began. All six artefacts were video; none was audio-only or still imagery. All six took a political subject, and five of the six concerned presidential succession, either a claimed candidacy, a claimed withdrawal, or a claimed rejection of a previous electoral choice. Five originated on TikTok and subsequently migrated to Facebook; one originated on X. Documented reach ranged from approximately 60,000 to 2.6 million views per incident. Captioning was predominantly Kiswahili or sheng, while every published correction was in English. Response latency, where it could be established, ran to several days between peak circulation and published verdict. Of the six incidents, one involved a documented response by an institutional actor; in the remaining five, the responding party was a fact-checking organization.

Four themes emerged from the second, inductive coding pass. The first addresses the first research question; the remaining three address the second.

Deepfakes in Kenya are provenance-parasitic

Across all six incidents, the synthetic artefact was not generated wholesale. In every case it was grafted onto authentic material, and in every case it borrowed the visual grammar of an institution entitled to be believed. In C1 the fabricated announcement was constructed over archival footage from the subject's presidency, the speaker positioned at a presidential lectern. In C4 the clip carried the logo of a private national broadcaster, implying that the station had published it. In C5 the underlying footage was traced to a party delegates' conference held on 26 September 2025. In C2 the artifact adopted the format of an international news analysis segment, complete with the identity of a recognized foreign correspondent.

This has a direct methodological consequence for verification, visible in how the fact-checkers actually worked. The decisive cues cited were rarely technical. In C1 the verdict turned on the colour of the lectern, blue having been used during the subject's presidency and yellow by the incumbent, and on the absence of any credible news record of such a speech. In C5, the analysis rested on reverse image search to establish the date and setting of the source footage, followed by a close comparison of a specific hand gesture. Algorithmic detection was used, but as corroboration rather than as the primary route to judgement; in C6 a detection tool returned a 99.6 percent likelihood of synthetic content, confirming a conclusion already reached contextually.

Provenance parasitism thus produces a paradox with practical implications. The artefact's persuasive force derives from its attachment to authentic material, and that same attachment is what renders it refutable, since the authentic original can be located and compared. The refutation, however, requires archival access, platform familiarity and time, resources concentrated in fact-checking organizations rather than in the communication offices of the institutions targeted.

The institutional correction deficit

In five of the six incidents, the corrective act was performed by a third-party fact-checking organization. In none of those five did the targeted principal, their communications office, or any government communications authority issue a documented rebuttal during the period between the fabrication's circulation and the fact-checker's verdict.

The absence is most consequential in C3. Synthetic resignation videos circulated during an acutely volatile political moment, reaching over 1.4 million views on a single platform, and the correction of record was published by a fact-checking organization. Coded against SCCT, this is a textbook victim-cluster crisis in which the recommended strategies of denial and clarification were available at negligible cost and were not deployed by the affected institution. The strategic vacuum was filled by an actor with no stake in the principal's reputation and no capacity to reach the audience that had seen the fabrication.

This finding recasts the problem. The Kenyan difficulty in 2025 was not that institutions responded badly to deepfakes. It is that, on the documented public record, they largely did not respond at all, and the response function was effectively outsourced to civil society.

Institutions function as vectors, not only as targets

C2 inverts the assumed structure of a deepfake crisis and is, for that reason, the corpus's most analytically productive case. A serving Principal Secretary for Foreign Affairs posted to his official account a synthetic video styled as international news analysis, presenting it as independent validation of Kenya's diplomatic conduct in the Sudan conflict and framing critics as fake analysts. The material was flagged by users, the post was deleted, and the following day the Principal Secretary issued a public apology acknowledging that he had inadvertently posted AI-generated content, thanking those who had flagged it, and announcing forward-looking measures including ministry work with technology partners on watermarking of synthetic content and the development of a training initiative on artificial intelligence for the diplomatic service (Owino, 2025).

Two observations follow. First, the vector was not a peripheral actor but a senior official using an official channel, which means institutional communication capacity is not merely inadequate to counter synthetic media but is itself a distribution surface for it. Second, this is the only incident in the corpus in which an institutional actor executed a complete crisis response, and the strategy deployed corresponds to none of the victim-cluster prescriptions. It combines acknowledgment, apology, credit to the correcting public, and announced corrective action, which is the structure of renewal rather than of denial. The corpus's single instance of effective institutional crisis communication about a deepfake occurred where the institution was culpable rather than victimized.

Corrections were channel-asymmetric

Across the corpus the fabrication and its correction inhabited different media environments. Fabrications originated predominantly on TikTok, migrated to Facebook and, by report, into closed messaging spaces, were captioned in Kiswahili and sheng, and took the form of short vertical video. Corrections were published as long-form English prose on fact-checker websites, subsequently syndicated to news aggregators and, through platform partnerships, applied as labels to some identified copies on Facebook and Instagram.

The mismatch is fourfold: platform, language, modality and register. In C6 the fabrication accumulated 2.6 million Facebook views; the correction took the form of an article. Only the platform-partnership labelling mechanism operates within the environment of circulation, and it applies to identified copies on two platforms, not to the originating platform, not to re-uploads, and not to closed channels, which Mudavadi et al. (2025) identify as a primary site of Kenyan misinformation encounter. Media Ecology Theory anticipates precisely this failure: a message formulated for the environment of the report cannot function in the environment of the feed.

Discussion

The attribution assumption does not hold

SCCT correctly classifies the Kenyan incidents. Its prescriptions, however, rest on an assumption the corpus does not support. The theory holds that responsibility attribution is the contested variable and that lowering attributed responsibility restores reputational position. In a deepfake crisis the contested variable is prior

to attribution: it is whether the event occurred. A denial in this setting is not a claim about responsibility but a claim about reality, and it is advanced by the party with the greatest interest in the outcome. In a low-trust environment such a claim is structurally weak, since the speaker's interest is precisely what is discounted.

This suggests amendment rather than rejection. SCCT's victim-cluster strategies should be understood as necessary but not self-sufficient in synthetic-media crises: denial must be *evidentially instantiated*, through production of the authentic original, and *relationally instantiated*, through relay by parties without a stake in the outcome. The corpus indicates that Kenyan fact-checkers already perform the second function; what is absent is any institutional practice of deliberately enlisting them rather than relying on them incidentally.

The counter-frame arrived late, or not at all

Framing Theory's expectation of a first-frame advantage is borne out, and compounded by the response latency documented in Section 5.2. Where the counter-frame is published days after the initial frame, in a different language, on a different platform, the conditions for displacement identified by Entman (1993) are simply not met. The corpus contains no instance of anticipatory framing, that is, of an institution establishing a general cautionary frame before an incident. This is notable because the prebunking literature indicates that such forewarning reduces susceptibility (Lewandowsky & van der Linden, 2021), and because Wack and Prochaska (2026) find that interpretive commentary surrounding a Kenyan deepfake conditions its effect, implying that the frame within which audiences encounter synthetic content is a genuine and available lever.

The liar's dividend as a foreseeable second-order risk

A further implication follows from the pattern documented in Section 5.2, though it is prospective rather than observed within the corpus. Where the routine institutional response to a fabrication is silence, and where fact-checking organizations become the recognized arbiters of authenticity, the public learns that the authenticity of any politically consequential recording is an open question awaiting external adjudication. This is the precondition for the liar's dividend that Chesney and Citron (2019) identify: once audiences accept that convincing fabrication is ordinary, a denial becomes available to any actor confronted with an authentic recording, and its plausibility no longer depends on evidence.

The corpus does not contain a documented instance of a Kenyan actor dismissing genuine material as synthetic, and this paper makes no claim that one occurred. What the corpus does establish is the enabling condition. Rini (2020) argues that the epistemic backstop is lost not when fabrication becomes possible, but when it becomes expected. On the evidence of 2025, Kenyan publics have been given ample instruction in that expectation and correspondingly little instruction in how authenticity is established. This is a strong argument for the anticipatory and aftercare components of the framework proposed below: institutions that build a public record of what is true retain a resource that institutions that respond only reactively do not have.

Channel symmetry as an operational requirement

Media Ecology Theory reframes the response deficit as an environmental mismatch rather than a resourcing failure. The practical implication is a determinate rule: the correction must occupy the same platform, language, modality, and register as the fabrication. This is not a stylistic preference but a condition of reaching the affected public. An institution whose crisis response capability consists of a press statement and a broadcast interview is, in the environment where deepfakes circulate, without a response capability at all.

Verification as a relational accomplishment

The three theories, taken together, explain the observed failures but do not fully account for the corpus's one success. C2 worked not because the response was fast, evidenced or well-framed, though it was

reasonably fast, but because the responding actor accepted the correction offered by the public, credited it, and converted the incident into a commitment. The audience was positioned as a party to verification rather than as its recipient. This is the structure Sawubona describes: recognition of the other as co-constituting what is known (Christians, 2004; Kamwangamalu, 1999).

It also matches how Kenyan audiences appear actually to verify. Where institutional trust is low, and circulation occurs in closed relational networks, a claim is assessed by reference to the person who forwarded it. Institutional crisis communication that ignores this operates on a model of credibility the environment has already abandoned. The framework proposed below takes it as the design principle.

The Relational Verification Framework

The Relational Verification Framework (RVF) is proposed as a context-specific model for public relations responses to deepfake crises in low-institutional-trust, high-circulation environments. It comprises four components arranged as a cycle rather than a sequence, together with one cross-cutting condition.

Component 1: Anticipatory framing. Before any incident, the organization establishes a standing cautionary frame and a provenance baseline. Operationally this means periodic public communication that synthetic impersonation exists and that official communication carries identifiable markers; cryptographic signing or watermarking of official audiovisual output, an approach already publicly endorsed within Kenya's foreign ministry following C2; and a maintained archive of authentic footage indexed by date and event, since Section 5.1 establishes that the authentic original is the decisive refutation asset.

Component 2: Contextual attribution. On detection, verification proceeds through contextual cues first and algorithmic detection second, in the sequence the corpus shows fact-checkers actually using: locate the source footage, establish its date and setting, identify appropriated credibility markers, and only then corroborate with detection tooling. This requires a named cross-functional team, a defined verification window, and a standing relationship with a fact-checking organization established before it is needed.

Component 3: Relational amplification. The correction is routed through the relational networks that carry the fabrication rather than broadcast from the institutional center. In the Kenyan setting, this means equipping trusted relays, including faith and community leaders, vernacular radio presenters, county communication officers and platform creators with existing audiences, with a short, evidence-based, vernacular correction they can carry in their own voice. This is the component with no counterpart in SCCT, and it follows directly from the relational conception of credibility set out in Section 3.4.

Component 4: Trust aftercare. The response does not terminate with the correction. Following the pattern of C2, it closes with acknowledgment of the public's role in detection, an explicit commitment to corrective action, and a subsequent report against that commitment. This addresses the residual uncertainty that Vaccari and Chadwick (2020) identify as the durable harm, and it converts each incident into an increment of the trust on which the next response will draw.

Cross-cutting condition: channel symmetry. Every component is executed in the platform, language, modality and register of the fabrication. A correction that fails this condition may be accurate, timely and well-evidenced, and will still not reach the public that saw the fake.

Operationally, the components map onto distinct time horizons, and the framework fails if they are treated as a single sequence initiated at the point of crisis. Anticipatory framing is continuous and must be in place before any incident, since neither provenance markers nor a footage archive can be constructed under time pressure. Contextual attribution belongs to the first hours and should be governed by a stated internal deadline, since the corpus indicates that the interval between peak circulation and published verdict is where reputational damage is incurred. Relational amplification belongs to the same day and depends on relationships cultivated in advance: a trusted relay approached for the first time during a crisis is an unfamiliar institution making an unverifiable request. Trust aftercare unfolds over subsequent weeks and

is the component most readily abandoned once immediate pressure subsides, which is precisely why it should be assigned to a named owner with a reporting date.

The framework's central claim is that its four components are interdependent rather than optional. Anticipatory framing without relational amplification produces warnings nobody relays; contextual attribution without trust aftercare produces a correct rebuttal that leaves residual doubt intact; relational amplification without prior provenance infrastructure asks trusted relays to vouch for a claim they cannot themselves verify.

Limitations and Directions for Future Research

Three limitations bound these conclusions. First, the corpus is restricted to incidents that fact-checking organizations selected for verification, which introduces a selection bias toward high-visibility political content and systematically excludes deepfakes circulating in closed channels, in corporate contexts, or against private individuals. Second, absence of a documented institutional response is not proof of absence: responses issued through closed channels, private platform escalation or informal newsroom contact would not appear in the public record, and the finding in Section 5.2 should be read as concerning the *public* response record. Third, the Relational Verification Framework is derived analytically from six cases and has not been empirically tested.

Three lines of further work follow. The framework should be validated through practitioner interviews with Kenyan communication professionals, conducted under institutional ethics clearance, to determine whether the absence of documented response reflects inaction or invisibility. Experimental testing of relational amplification against centralized correction, extending the design used by Wack and Prochaska (2026), would assess whether the model's central mechanism performs as predicted. Finally, comparative application across East African settings with differing trust profiles would test the framework's transferability ahead of the region's upcoming electoral cycles.

Conclusion

This study set out to establish how deepfakes circulated in Kenya in 2025, how actors responded, whether established theory accounts for those responses, and what model might better serve practice. The record shows synthetic artifacts built parasitically on authentic footage and institutional insignia; corrections performed almost entirely by civil society rather than by the institutions concerned; institutional actors serving as both vectors and targets; and corrections systematically mismatched to the environments in which fabrications circulated.

The theoretical implication is specific rather than sweeping. Situational Crisis Communication Theory retains its classificatory power but requires amendment when the contested question is what occurred rather than who is responsible, because in that case a truthful denial from an interested party is structurally weak. Framing Theory and Media Ecology Theory together explain why late, centralized, English-language corrections fail. None of the three provides an account of how credibility is conferred where institutional trust is low. At that point, the Sawubona conception of verification as relational accomplishment becomes analytically necessary rather than merely culturally apt.

The practical stakes are immediate. Kenya's 2027 electoral cycle has already begun to generate synthetic content, and the corpus analyzed here documents an institutional response capability that, on the public record, was scarcely exercised in 2025. The Relational Verification Framework is offered as a working protocol rather than a settled model: anticipate, attribute contextually, amplify relationally, and repair trust afterward, executing each in the language, on the platform, and in the register where the falsehood actually lives. Whether Kenyan institutions build that capability in the time available to them is not a research question. It is a decision.

References

African Union. (2024). Continental artificial intelligence strategy: Harnessing AI for Africa's development and prosperity. African Union Commission.

Africa Check. (2025a, April 4). Ignore deepfake video claiming former Kenyan president Uhuru Kenyatta to run for president in 2027. https://africacheck.org/fact-checks/meta-programme-fact-checks/ignore-deepfake-video-claiming-former-kenyan-president-uhuru

Africa Check. (2025b, July 7). Ignore AI-generated videos used to falsely claim that Kenyan president Ruto has resigned. https://africacheck.org/fact-checks/meta-programme-fact-checks/ignore-ai-generated-videos-used-falsely-claim-kenyan

Africa Check. (2025c, December 3). Ignore manipulated video showing former Kenyan president Uhuru Kenyatta appearing to hint at 2027 presidential bid. https://africacheck.org/fact-checks/meta-programme-fact-checks/ignore-manipulated-video-showing-former-kenyan-president

Africa Check. (2025d, December 4). Ignore manipulated video of former Kenyan president Kenyatta calling 2022 election of president Ruto a mistake. https://africacheck.org/fact-checks/meta-programme-fact-checks/ignore-manipulated-video-former-kenyan-president-kenyatta

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

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753-1820. https://doi.org/10.15779/Z38RV0D73H

Christians, C. G. (2004). Ubuntu and communitarianism in media ethics. Ecquid Novi: African Journalism Studies, 25(2), 235-256.

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. Corporate Reputation Review, 10(3), 163-176. https://doi.org/10.1057/palgrave.crr.1550049

Coombs, W. T. (2015). Ongoing crisis communication: Planning, managing, and responding (4th ed.). Sage.

Coombs, W. T., & Holladay, S. J. (2012). The paracrisis: The challenges created by publicly managing crisis prevention. Public Relations Review, 38(3), 408-415. https://doi.org/10.1016/j.pubrev.2012.04.004

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Sage.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51-58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Fink, S. (1986). Crisis management: Planning for the inevitable. AMACOM.

Frandsen, F., & Johansen, W. (2017). Organizational crisis communication: A multivocal approach. Sage.

Jin, Y., Liu, B. F., & Austin, L. L. (2014). Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on publics' crisis responses. Communication Research, 41(1), 74-94. https://doi.org/10.1177/0093650211423918

Kamwangamalu, N. M. (1999). Ubuntu in South Africa: A sociolinguistic perspective to a pan-African concept. Critical Arts, 13(2), 24-41.

Karanja, J. M., & Nzau, T. (2026). The emergence of AI deepfakes and their impact on strategic crisis communication: Practitioners' perspectives from Kenya. Journal of Contingencies and Crisis Management, 34(1), e70140. https://doi.org/10.1111/1468-5973.70140

Kenya. (2018). The Computer Misuse and Cybercrimes Act, 2018. Government Printer.

Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. European Review of Social Psychology, 32(2), 348-384. https://doi.org/10.1080/10463283.2021.1876983

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage.

Mbigi, L. (1997). Ubuntu: The African dream in management. Knowledge Resources.

McIntyre, L. (2018). Post-truth. MIT Press.

Mudavadi, K. C., & Madrid-Morales, D. (2024). Countering political disinformation. In The Routledge companion to digital journalism studies (pp. 450-459). Routledge.

Mudavadi, K. C., Tully, M., & Lomoywara, D. B. (2025). Exploring Kenyans' interactions with misinformation on WhatsApp. Mobile Media & Communication, 13(1), 69-87.

Nation Africa. (2025, December 29). Deepfakes, hashtags and hate: The 2025 online disinformation threatening Kenya's 2027 polls. https://nation.africa/kenya/news/deepfakes-hashtags-and-hate-the-2025-online-disinformation-threatening-kenya-s-2027-polls--5310674

Owino, W. (2025, February 21). I'm sorry, Foreign Affairs PS Sing'oei says over deep fake video. The Standard. https://www.standardmedia.co.ke/national/article/2001512306/im-sorry-foreign-affairs-ps-singoei-says-over-deep-fake-video

Patton, M. Q. (2015). Qualitative research and evaluation methods: Integrating theory and practice (4th ed.). Sage.

PesaCheck. (2025, November 20). Altered: This video of Uhuru Kenyatta cautioning Kenyans on 2027 elections is doctored. https://medium.com/pesa-check

Postman, N. (1970). The reformed English curriculum. In A. C. Eurich (Ed.), High school 1980: The shape of the future in American secondary education (pp. 160-168). Pitman.

Rini, R. (2020). Deepfakes and the epistemic backstop. Philosophers' Imprint, 20(24), 1-16. https://doi.org/10.3998/phimp.115

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. University of Illinois Press.

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

Veil, S. R., Buehner, T., & Palenchar, M. J. (2011). A work-in-process literature review: Incorporating social media in risk and crisis communication. Journal of Contingencies and Crisis Management, 19(2), 110-122. https://doi.org/10.1111/j.1468-5973.2011.00639.x

Wack, M., & Prochaska, S. (2026). Making sense of AI-generated disinformation: How audience interpretations influence the impact of deepfakes in Kenya. Social Media + Society. Advance online publication. https://doi.org/10.1177/20563051261462092

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy. Council of Europe.