Journal of Visual and Performing Arts
Vol. 4 No. 1 (2026): ISSN (Online): 2957-5990
DOI: https://doi.org/10.58721/cehhyv68
Practising Algorithmic Authenticity in East African Public Relations: A Framework Grounded in Kenyan Cases

Practising Algorithmic Authenticity in East African Public Relations: A Framework Grounded in Kenyan Cases

Cynthia Shimanyula
Kabarak University, Kenya

Article History
Received: 2026-03-20
Revised: 2026-08-30
Accepted: 2026-09-01
Published: 2026-09-08

Keywords
Africa
Algorithmic authenticity
Ethics
Justice

How to cite:
Shimanyula, C. (2026). Practising Algorithmic Authenticity in East African Public Relations: A Framework Grounded in Kenyan Cases. *Journal of Visual and Performing Arts, 4*(1), 70-83.

Abstract
Generative artificial intelligence has moved rapidly from novelty to everyday instrument in East African public relations, strategic communication, and institutional visual production. Adoption is largely ungoverned, and the tools carry the aesthetic, linguistic, and cultural defaults of the Global North data on which they are trained — an issue of algorithmic justice with direct consequences for media integrity in the region. This paper reconceptualises authenticity, long treated as the currency of public trust in PR, as an ongoing practice rather than a fixed property a message possesses or lacks: a set of deliberate ethical decisions practitioners make each time they engage an AI tool. Drawing on sixteen years of PR practice and teaching in Kenya, the paper proposes a four-dimensional framework — disclosure, cultural-linguistic fidelity, human accountability, and proportionality — and illustrates it through Kenya as an information-rich empirical entry point into a broader East African and Global South dynamic. Two documented Kenyan cases (Posta Kenya's AI-generated customer-service image and the Kenya Urban Roads Authority's culturally implausible road imagery) anchor the analysis, read through visual social semiotics and a decolonial lens. A corroborating case — the US State Department's AI-generated map at the AIDS 2026 conference, which misrepresented Uganda and other African countries — extends the framework beyond the region's own institutions. Faith-based communication is noted as a setting where the framework's questions sharpen further, contributing a locally grounded, teachable framework for practitioners, educators, and institutions across East Africa.

Introduction
Across East Africa, generative artificial intelligence (AI) has moved rapidly from novelty to everyday instrument in public relations (PR), strategic communication, and institutional visual production. Practitioners now use generative tools to draft copy, monitor media, produce visual content, and manage audience engagement, drawn by the promise of speed and cost-saving in resource-constrained communication environments. Adoption has been swift and largely ungoverned: few organisations in the region have formal policies on when and how such tools may be used, and practitioners frequently reach for them without institutional guidance. This paper asks what this shift means for the value on which the profession most depends, and argues that answering that question well requires rethinking what authenticity is when the tools of communication, including its images, are themselves generative AI systems.

Authenticity has long been treated as the currency of public trust in PR. Yet contemporary scholarship establishes that authenticity is not an intrinsic property of a message; it is a perception that publics form in context, weighing what they see against what they know of the communicator (Lim & Jiang, 2022). This distinction matters greatly once generative tools enter the production of communication artifacts, and it matters most acutely for the visual artifact, whose cultural and representational failures are legible at a glance. If authenticity is a judgement public make in context, then introducing a system that generates text or imagery is not a neutral gain in efficiency; it is an intervention into the very grounds on which that judgement is made.

This question is sharpened in the East African setting by the provenance of the tools themselves. Leading generative systems are trained overwhelmingly on data drawn from the Global North, and they carry the aesthetic, linguistic, and cultural defaults of that data. Critical scholarship characterises this dynamic as an algorithmic colonisation of Africa (Birhane, 2020) and positions decolonial theory as a necessary lens for anticipating such harms (Mohamed et al., 2020). These pressures are shared across the wider Global South, but they take a specific form in East Africa, where fact-checking infrastructure and institutional guardrails are comparatively thin.

The stakes become most publicly visible in generative visual content, where cultural and representational failures are legible at a glance. This paper takes as its illustrative entry point, two documented episodes in which Kenyan public institutions released AI-generated images that failed before their own audiences. These Kenyan cases are treated not as the limit of the argument but as an information-rich window into a regional and Global-South dynamic. The framework the paper develops is illustrated through visual cases and grounded in visual social semiotics, but is modality-agnostic and applies equally to generated text, audio, and multimodal outputs.

Literature review

The framework builds on four established bodies of theory, each illuminating part of the problem while leaving a gap the framework addresses.

Authenticity and trust in strategic and visual communication

Authenticity has long been treated in PR scholarship as foundational to the organization–public relationship and to the trust on which that relationship depends. Recent work continues to demonstrate that perceived authenticity in organisational communication shapes relationship quality, credibility, and reputational outcomes, and that audiences increasingly demand communication they read as genuine rather than merely instrumental (Lim & Jiang, 2022). This literature establishes that authenticity is relative, contextual, and continually assessed by publics against the organisation's conduct and claims. Visual-communication scholarship, drawing on visual social semiotics (Kress & van Leeuwen, 2021), adds that images are read as meaning-making artifacts through their represented participants, modality, and compositional cues, so that visual authenticity is judged through the interpersonal relationship the image constructs with the viewer. What neither literature yet addresses is what authenticity requires when the communicator is no longer solely human, and when the tools themselves carry cultural assumptions foreign to the audience. This strand grounds the framework as a whole and the dimension of disclosure in particular.

The ethics of transparency and moral accountability

A second strand concerns who can be held responsible for a communicative act. Communication-ethics scholarship treats transparency and honesty as core professional obligations and locates moral agency in a human actor who can answer for what is produced. Matthias (2004) influentially identified a responsibility gap that arises when the behaviour of learning systems can no longer be fully predicted or controlled by their operators, such that traditional ascriptions of responsibility no longer

straightforwardly apply. The argument applies directly to generative AI in PR and visual communication: when an artifact is substantially produced by a model whose outputs its user did not author and cannot fully anticipate, the question of who is accountable for that artifact becomes pressing. This strand grounds the dimensions of disclosure, especially human accountability.

Algorithmic justice and decolonial approaches to technology

The third strand is the most theoretically alive for an East African argument, and it grounds the dimension of cultural-linguistic fidelity. Birhane (2020) argues that the large-scale deployment of AI systems trained on and controlled from the Global North amounts to an algorithmic colonisation of Africa. Mohamed et al. (2020) show how decolonial theory can serve as a sociotechnical foresight tool for identifying coloniality in AI. Applied specifically to African strategic communication, Ooko (2025) argues for a decolonial framework for communicating AI in African public service. Empirical audits confirm the concern in the visual domain: Alenichev et al. (2023) found a leading image tool structurally incapable of depicting Black African professionals caring for white patients, and Bianchi et al. (2023) documented text-to-image systems rendering African subjects through markers of underdevelopment. What this literature does not yet offer is an operational account of what culturally faithful communication looks like when the default outputs of generative tools flatten or misrepresent local languages, registers, and imagery.

Empirical work published in 2025 and 2026 sharpens the picture in ways directly relevant to the East African case, and to the Kenyan cases analysed below in particular. AlDahoul et al. (2025) document, in a large-scale audit of image outputs, a phenomenon of racial homogenisation in which members of a given race are depicted as strikingly similar to one another, and also demonstrate that considered debiasing interventions materially reduce those effects. Weinmann et al. (2026), in a preregistered content analysis across DALL·E, Midjourney, and Stable Diffusion, confirm that both representational and presentational biases persist across the leading platforms though their magnitude varies by prompt and by tool. Yang (2025) extends the evidence beyond text-to-image to image-to-image generation, finding that images of people of colour are reproduced with substantially lower racial accuracy than images of White people even when the subject is unmistakably present in the input. A PRISMA-based systematic review by Moletsane (2026) synthesises this pattern across the recent literature and concludes that the limited representation of African populations, languages, and contexts in training data contributes directly to algorithmic bias. Read together, these studies document that the training-data problem is empirically real and structural, and also that considered intervention on the model and on the prompt can measurably mitigate it — a finding that grounds the framework's insistence on cultural-linguistic fidelity as an active practitioner discipline rather than a hope that the tools will fix themselves.

It bears emphasising that the representational patterns generative systems reproduce are not new; they are the automation and scaling of an older media grammar. Scholarship on the representation of Africa in Western media documents, over many decades, the recurring depiction of the continent through a narrow repertoire of tropes; poverty, disease, wildlife, an undifferentiated 'African' aesthetic, and the white saviour positioned against a suffering Black subject; patterns rooted in colonial iconography and reinforced through global-health, news, and NGO imagery. The visual failures documented in the recent generative literature, from the structural inability of a leading tool to depict Black African professionals caring for white patients (Alenichev et al., 2023) to the rendering of African subjects through markers of underdevelopment (Bianchi et al., 2023), belong recognisably to that longer lineage rather than constituting a fresh problem invented by machine learning. What generative systems change is scale, speed, and the sites of reproduction: pre-AI stereotypes travelled through editorial workflows in which at least some human eyes intervened; algorithmic ones travel

through automated pipelines at a rate no editorial function catch. The Kenyan cases analysed below should therefore be read as the machine-scaled continuation of an age-long visual grammar, not as its arrival, and this continuity is what makes the framework's fidelity dimension both urgent and answerable to a tradition of critique that predates AI.

Proportionality and risk in technology ethics

The fourth strand concerns matching the depth of a technology's involvement to the stakes of its use. Frameworks for the ethics of AI commonly organise themselves around a balance of opportunity against risk (Floridi et al., 2018). Embedded in this literature is a proportionality logic: caution, human oversight, and justification should rise with the stakes of the application. This logic is well developed at societal-governance level but rarely translated to the individual communicative act. The paper's dimension of proportionality adapts this governance-level insight to the practitioner's everyday judgement. This adaptation must, however, answer Mittelstadt (2019), who argues that principle-based approaches to AI ethics may fail to guide practice because AI work lacks common aims and proven methods for translating principles into action. The paper responds by grounding each dimension in concrete, observable decisions and pairing the framework with disclosure norms and cultural quality-assurance routines.

The gap and the present contribution

Taken together, these strands establish that authenticity is a contextual perception on which trust depends, that generative systems disrupt settled ascriptions of accountability, that AI tools carry culturally and colonially inflected assumptions that bear directly on communication in Africa, and that ethical use should be proportional to the stakes. What none provides, alone or together, is an integrated, practitioner-facing account of how to act authentically when the tools of communication are themselves generative AI systems, the training data is predominantly of the Global North, and the audience is East African. This paper offers such an account.

Method

This section sets out the study's research design and paradigm, the procedure used to develop the four-dimensional framework, the documentary case analysis that grounds it, and the criteria applied to analytical rigour, reflexivity, and ethics.

Research approach and paradigm

This study employs a conceptual-analytical research design augmented by illustrative documentary case analysis. Its primary contribution is theoretical: it reconceptualises authenticity in AI-mediated PR and visual communication and advances a four-dimensional framework of algorithmic authenticity. Consistent with Jaakkola's (2020) typology of conceptual articles, the design corresponds most closely to the theory-synthesis and theory-adaptation modes. The paper does not seek to test the framework statistically; rather, it develops the framework and demonstrates its analytical utility, an aim for which conceptual and qualitative-interpretive designs are appropriate (Rocco et al., 2022). Among the three research paradigms most commonly distinguished in social inquiry — positivist, relativist, and pragmatic — the study is positioned within a relativist (interpretivist) paradigm with a critical orientation: meaning is understood as constructed through interpretation in context rather than discovered as an objective fact, and interpretation is exercised with attention to the power relations that shape which meanings are produced and by whom.

Development of the conceptual framework

The four-dimensional framework was constructed through a structured synthesis of three bodies of scholarship, integrated with practitioner insight. The first comprises established PR and communication-ethics scholarship and the professional codes that formalise it, from which disclosure

and human accountability are derived. The second comprises the emerging literature on AI ethics and the governance of generative systems, from which the concept of proportionality is developed. The third comprises scholarship on algorithmic justice, algorithmic bias, and decolonial approaches to communication and technology, from which cultural-linguistic fidelity is drawn and given its East African inflection. These constructs were integrated with the author's insight as a practitioner and educator with sixteen years of mixed PR and communication-teaching experience in Kenya, engaged as experiential rather than systematically documented knowledge.

The framework is anchored in four established theoretical framings, each contributing one dimension. Communication ethics and the PR literature on authenticity and impression management (Lim & Jiang, 2022) ground disclosure. Matthias's (2004) account of the responsibility gap, developed within the ethics of learning systems, grounds human accountability. Decolonial theory and algorithmic-justice scholarship (Birhane, 2020; Mohamed et al., 2020; Ooko, 2025) ground cultural-linguistic fidelity, with visual social semiotics (Kress & van Leeuwen, 2021) supplying the reading protocol for images. The principlist and proportionality strands of AI ethics (Floridi et al., 2018; Mittelstadt, 2019) ground the principle of proportionality. These framings are integrated with the author's sixteen years of Kenyan PR practice and communication teaching and are treated as experiential rather than as systematically documented knowledge.

The four dimensions are analytically distinct but empirically interlocked. Disclosure and human accountability form the honesty-and-answerability axis: the first commits the communicator to naming AI involvement openly, the second commits a named person to answering for the artifact regardless of the tool that produced it. Cultural-linguistic fidelity and proportionality form the fit-and-restraint axis: the first commits the communicator to outputs that speak to and from the local context, the second commits the communicator to matching the depth of AI involvement to the stakes of the communication. A failure on any single dimension can be corrected; the cases analysed below fail on all four together, and the failures compound because each dimension both draws on and reinforces the others. Figure 1 illustrates these interrelations diagrammatically.

Journal of Visual and Performing Arts
Vol. 4 No. 1 (2026): ISSN (Online): 2957-5990
DOI: https://doi.org/10.58721/cehhyv68
Practising Algorithmic Authenticity in East African Public Relations: A Framework Grounded in Kenyan Cases

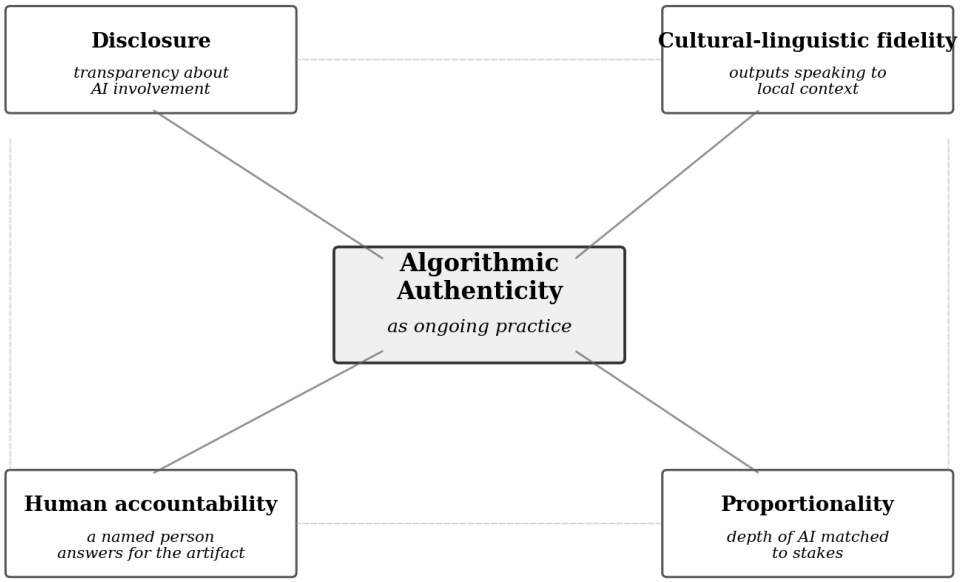

Figure 1: The four dimensions of algorithmic authenticity and their interrelations.

Documentary case analysis and visual reading

To ground the framework, the study incorporates documentary analysis of two publicly reported cases of AI-generated visual content in Kenyan institutional communication. Documentary analysis is an established qualitative method (Bowen, 2009; Morgan, 2022), and the case orientation follows the interpretive case-study tradition (Stake, 1995; Yin, 2018). Cases were chosen through purposive, information-oriented selection on the logic that a well-chosen case yields more analytical leverage for theory development than a large but undifferentiated sample (Flyvbjerg, 2006). Three criteria were applied: documented use of generative AI in outward-facing strategic or brand communication by an identifiable organisation; documentation by at least two independent, credible sources; and Kenyan setting so that questions of cultural and linguistic fidelity would be analytically legible.

The two cases are engaged through secondary sources: the images as they were publicly circulated, published news reports, practitioner commentary, and documented public reactions on visible platforms. The study has no primary access to the organisations concerned, to the internal decisions behind the AI adoption, or to the artifacts as they existed in the production pipeline. All claims are accordingly claims about the public record of these episodes as it can be reconstructed from that record. Documentary analysis is designed for exactly this kind of reading.

Two cases satisfying these criteria were selected. The primary case is the Postal Corporation of Kenya (Posta Kenya), which circulated an AI-generated customer-service image whose figures displayed anatomical anomalies and whose depicted Kenyan staff were rendered, in part, as white, before the image was withdrawn without public explanation. The corroborating case is the Kenya Urban Roads Authority (KURA), which circulated AI-generated imagery promoting national road connectivity that publics judged culturally implausible. Because the artifacts are visual, the analysis was informed by

the principles of visual social semiotics (Kress & van Leeuwen, 2021), reading each image for its represented participants, modality, compositional cues, and the interpersonal relationship it constructs with its viewer. A consistent analytical boundary was maintained: the study analyses the artifacts and their public traces, not the internal decision-making of the organisations concerned.

Analytical rigour, reflexivity and ethics

Trustworthiness was pursued using the established criteria for qualitative inquiry (Lincoln & Guba, 1985): source triangulation across multiple independent reports per case, an auditable trail linking each interpretive claim to a dated source, thick contextual description, and a transparent, replicable procedure applying the same four-dimension protocol to each case. The author writes as both a Kenyan PR practitioner and a communication educator, and this dual standpoint is treated as an analytical resource rather than a neutral vantage point (Sibbald et al., 2025); the risk of confirmation bias was mitigated by pre-specifying analytical criteria and grounding every claim in the corroborated public record. The study analyses publicly available organisational communications and publicly visible reactions; it involves no human participants, no private data, and its critique is directed at practices and structures rather than at identifiable individuals.

Results

This section applies the four dimensions of algorithmic authenticity to the two documented Kenyan cases, reading each artifact and its public reception against each dimension in turn. In the primary case, Posta Kenya published an AI-generated promotional image on the platform X, featuring four smiling customer-service agents wearing headsets, accompanied by a caption inviting the public to contact its care team (Gachie, 2024; The Kenya Times, 2024). In the corroborating case, KURA published AI-generated images promoting national road connectivity across the country's counties (Citizen Digital, 2024; AIpots, 2024).

Disclosure

In neither of the two cases did the institution label its image as AI-generated; in both cases the public inferred the artifact's synthetic origin from visible anomalies rather than from any disclosure by the communicator (Gachie, 2024; Citizen Digital, 2024; AIpots, 2024). Disclosure is the honesty dimension of algorithmic authenticity, and its breach here is instructive. Because authenticity is a perception publics form in context (Lim & Jiang, 2022), the manner in which AI involvement comes to light matters as much as the fact of it. When provenance is undisclosed and then detected adversarially, by audiences catching a third hand or a misrendered face, the communicator is revealed as not having been transparent, and the trust relationship suffers a second injury beyond the visual error itself. Proactive disclosure would not have prevented the anomalies, but it would have changed the ethical texture of the episode from concealment-exposed to openness-with-imperfection.

Cultural-linguistic fidelity

The cultural-linguistic fidelity dimension reveals the most consequential and distinctly regional failure. In the Posta Kenya image, members of the public observed that two of the four supposedly Kenyan customer-service agents were rendered as white (Gachie, 2024); in the KURA images, audiences judged the depicted scenes culturally implausible, reading them as generic and dystopian rather than recognisably Kenyan (Citizen Digital, 2024; AIpots, 2024). These are not incidental glitches but the visible signature of training data drawn overwhelmingly from the Global North, whose demographic and aesthetic defaults surface unless actively resisted. This is precisely the dynamic that critical scholarship names as the algorithmic colonisation of Africa (Birhane, 2020), which decolonial approaches treat as foreseeable rather than accidental (Mohamed et al., 2020). That a Kenyan state corporation should, through an unexamined tool, depict its own workforce as partly white is

algorithmic injustice made concrete in everyday institutional communication, exactly the terrain that recent work on the decolonial communication of AI in African public service identifies as demanding attention (Ooko, 2025).

This argument requires a contemporary qualification. The most recent generation of image and multimodal models can, with careful prompting and when trained on more geographically inclusive data, produce African-facing outputs of markedly higher cultural fidelity than the systems available at the time of the cases analysed here. The empirical literature confirms both sides of this picture: recent audits document persistent representational and presentational biases across the leading platforms (AlDahoul et al., 2025; Weinmann et al., 2026; Yang, 2025) while also demonstrating that considered debiasing interventions and prompt discipline can materially reduce those biases (AlDahoul et al., 2025). The framework's implication is therefore not that the tools are irredeemably biased but that default outputs, unaudited and unprompted for local context, will tend to reproduce training-data defaults, and the practitioner burden is to interrupt that tendency deliberately.

Human accountability

Those artifacts displaying supernumerary hands, misrendered staff, and roads terminating without destination reached publication at all indicates the absence of meaningful human review before release, and in the primary case the institution withdrew the image after a short interval without offering any public account of what had occurred (Gachie, 2024; NTV Kenya, 2024). The human accountability dimension reads this as a failure of moral agency. Generative systems produce outputs their users did not author and cannot fully anticipate, opening the responsibility gap in which traditional ascriptions of answerability no longer hold (Matthias, 2004). Algorithmic authenticity closes that gap by insisting that a named human retain and exercise judgement such that someone can answer for the artifact. On both counts the cases fall short: no evident human judgement intercepted the flawed images before release, and the silent withdrawal left no identifiable person answering afterwards.

The point of the argument here is deliberately directed at the practitioner and the institution, not at the model. The images did not post themselves; an identifiable Kenyan communications officer, at an identifiable desk, in an identifiable institution, took the decision to release each artifact without cultural review, without prompt revision, and without any evident check against the audience the image would reach. That the tool made the output easy to produce is a fact about the tool; that the output reached publication is a fact about a person's judgement. A useful analogue is the familiar principle that a manufacturer is not held responsible for the criminal use another person makes of a product it made lawfully available: the responsibility travels with the user's decision, not with the tool's existence. The framework's insistence on human accountability is precisely that responsibility travels with the person who chose to release the artifact, and any adequate response to failures of the kind analysed here must begin there. The proper technical corollary is a demand on the wider system: the response to poorly localised outputs is to build, adapt, or fine-tune models on data that serve Kenyan communicative intent — a task Kenyan institutions and researchers can lead — rather than to accept the defaults as fixed. In this sense the tool question and the practitioner question meet: better tools reduce the cost of good practice, but they do not replace it.

Proportionality

Both artifacts were official communications from national public institutions—a state postal corporation and a statutory roads authority—addressed to the general public. Such communication carries high credibility stakes: it represents institutions on which citizens rely, and its authority depends on being taken seriously. The proportionality dimension holds that the depth of AI

involvement should be matched to those stakes (Floridi et al., 2018). Using unaudited generative imagery for official institutional communication inverts that logic, applying the least oversight precisely where the stakes are highest. The episodes sit within a broader shift in which leading Kenyan brands have moved to generative creative assets on cost grounds, a shift a prominent practitioner warned risked eroding the very Kenyan authenticity those brands depend upon (Kemibaro, 2024).

Cross-dimensional findings

Read across the four dimensions, the cases do not fail on one axis while succeeding on the others; they fail on all four at once, and the failures compound. An undisclosed image that misrepresents the local community, which no human judgement intercepted and for which no person answered, deployed in a high-stakes official channel, is a single integrated failure of authenticity-as-practice rather than four separate defects. Non-disclosure converts a technical imperfection into a breach of trust, confirming that transparency about AI involvement is constitutive of, not incidental to, perceived authenticity (Lim & Jiang, 2022). Cultural-linguistic fidelity is the most consequential and most distinctively East African dimension: the rendering of public-sector workers as partly white gives concrete, locally legible form to what has been theorised as algorithmic coloniality. Accountability failures are structural rather than incidental. And both cases are proportionality failures, since high-stakes official communication is precisely where unaudited generative shortcuts are least defensible.

A corroborating case beyond the Kenyan setting

The framework's diagnostic reach is worth testing on a case whose organisational locus lies outside Kenya but whose representational subject sits squarely within the East African region. In July 2026, at the AIDS 2026 conference in Rio de Janeiro, the head of the US President's Emergency Plan for AIDS Relief presented an official slide displaying an AI-generated map of Africa that mislabelled every country it identified; Uganda, Mozambique, Nigeria, Cameroon, Malawi, and Côte d'Ivoire and that a Reuters analysis subsequently confirmed carried an OpenAI watermark (Al Jazeera, 2026). The US State Department accepted responsibility, describing the slide as an unfortunate error produced by a staff member who had hastily modified the deck before the event, and apologised to attendees including its African partners (Al Jazeera, 2026).

Read through the four dimensions, the case fails on the same axes as the Kenyan ones and for structurally similar reasons. AI provenance was not disclosed; it was detected by an outside analyst rather than acknowledged by the communicator, converting a technical failure into a transparency failure. The cultural-linguistic fidelity failure is the paradigm instance of the coloniality dynamic the paper theorises: a Global North institution used an AI system to represent African countries and the system produced a map on which Nigeria appeared landlocked in the Sahara, Mozambique migrated to the Horn of Africa, and Uganda was rendered with a boundary bearing no relation to its territory — the mirror image of the Posta Kenya failure, in which a Kenyan institution used AI to depict its own workforce as partly white. No evident human review caught the slide before it reached a high-stakes international conference, and the apology attributed the error to an unnamed staff member rather than to any identifiable review process. Proportionality was catastrophically mismatched: an official multi-country health-funding presentation is precisely the setting where unaudited generative shortcuts are least defensible. Commenting on the incident, Moky Makura of Africa No Filter observed that the deeper question is what the AI was drawing on to get every country wrong (TRT Afrika, 2026), which is precisely the point the cultural-linguistic fidelity dimension names. The corroborating value of this case is that the four dimensions apply to Global North organisations whose AI-generated communication represents East African subjects, and that the failure pattern is constitutive of ungoverned generative-AI use in institutional communication about Africa wherever it occurs.

A counter-reading and the framework's limits

Analytical honesty requires testing whether the framework could be wrong. The cases were selected because they are vivid failures; a framework that only ever finds fault is not diagnostic but merely accusatory. The framework resists this because its dimensions are separable and can return positive as well as negative judgements: an artifact that is openly labelled, culturally faithful, humanly reviewed, and stakes-appropriate would pass all four. One might also argue that public backlash, not the framework, is doing the analytic work. This is partly true and a strength: the framework explains why the backlash was warranted and locates the breach precisely, distinguishing a fidelity failure from an accountability failure that a diffuse sense of "AI slop" conflates. Mittelstadt (2019) would caution that naming dimensions does not, by itself, change practice; only the professional adoption of disclosure and cultural-assurance routines could, which is why the framework is offered alongside the practical and pedagogical infrastructure set out in the Discussion.

The framework's positive returns are worth naming, as are its diagnostic ones. An artifact would pass all four dimensions where its producer had disclosed AI involvement openly at the point of release, had subjected the output to cultural review before publication and revised or replaced it where the review identified a fidelity problem, had named the human editor who authorised the release and remained available to answer for it, and had matched the depth of AI involvement to the stakes of the communication; light AI assistance for low-stakes internal messaging, tighter and human-centred production for high-stakes public artifacts. Emerging Kenyan practice is beginning to produce examples in this direction: agencies and in-house teams that publish AI-use disclosures on their websites, editorial teams that route AI-generated visuals through a named human reviewer before release, and public-service communicators experimenting with locally grounded prompt discipline informed by decolonial framings (Ooko, 2025). Regional policy movement is beginning to codify the disclosure dimension in particular: a draft Rwandan media policy circulated in 2023–2024 proposes mandatory labelling of AI-generated content across social media, news platforms, and search engines, offering an East African regulatory template for the disclosure norm the framework prescribes (Rwanda Inspirer, 2024). The prescriptive value of the framework rests as much on the practicality of these positive moves as on the diagnostic value of the negative cases.

Discussion

The analysis demonstrates that the loss of authenticity in each case was produced not by a fixed property of the institution but by a sequence of human decisions: to use the tool, not to audit its output, not to disclose its involvement, and not to answer for the result. This directly supports the paper's central reconceptualisation: authenticity in an AI-mediated environment is not a quality an institution statically possesses or lacks, but an ongoing practice enacted, or neglected, decision by decision. The framework performed its intended diagnostic work, locating and naming each failure precisely and distinguishing dimensions that a purely aesthetic account would have collapsed.

A further explanatory question follows: why do such publicly conspicuous errors occur in Kenyan institutional communication when they are rarer in otherwise comparable Global North settings? A plausible part of the answer lies at the organisational level. Where institutional consequences for reputational lapse, internal review, reassignment, resignation, and sometimes regulatory sanction are routine and expected, communicators anticipate those consequences and audit their outputs accordingly. Where such consequences are diffuse or absent, the ordinary discipline that a rigorous consequence architecture imposes is missing, and the space for reckless shortcuts widens. The failures documented here are therefore not only failures of individual judgement but symptoms of an escapism that thin consequence structures permit. The prescriptive counterpart is direct: algorithmic

authenticity is far likelier to be practised in institutions whose communicators know, routinely and credibly, that a botched artifact will be traced, named, and answered for.

Read against sixteen years of Kenyan PR practice and communication teaching, three patterns visible in the cases are familiar rather than surprising. First, cost pressure is the dominant driver of tool adoption in Kenyan communications teams, and the discipline of cultural review is the first thing squeezed when a campaign is due tomorrow. Second, responsibility for outward-facing content in many Kenyan institutions is distributed thinly and often informally, so that when an artifact fails publicly no single person is positioned, or willing, to answer for it. Third, in the classroom setting communication students arrive fluent in the tools long before they arrive fluent in the ethics of using them; teachable, compact frameworks are therefore of practical value, which is part of why the four dimensions have been kept few and deliberately learnable. These observations do not substitute for empirical evidence, and the paper does not treat them as such, but they shape which parts of the framework are pitched to the practitioner, which to the institution, and which to the educator.

Implications for practice, pedagogy, and policy

For practice, the framework translates into working habits: disclosing AI involvement rather than allowing publics to discover it; instituting cultural quality assurance so that outputs are checked against the community they address before release; ensuring a named person retains and exercises judgement over each artifact; and proportioning the depth of AI involvement to the stakes of the communication. For pedagogy, the framework is deliberately compact enough to be taught in a single session, equipping diploma and undergraduate communication students who will enter AI-saturated workplaces with a vocabulary and a decision-making procedure rather than merely a warning. For institutions developing AI-use policies, the four dimensions offer a ready checklist against which draft guidelines and individual campaigns can be assessed.

Boundary conditions

The framework is not universal. It is designed for intentional, outward-facing strategic and visual communication by an identifiable communicator who can be held to account; it is not a tool for adjudicating covert disinformation, user-generated content, or fully automated systems with no human in the loop. Its cultural-linguistic fidelity dimension presupposes a reasonably identifiable target community against which fidelity can be judged. Its proportionality dimension assumes the communicator has discretion over how much AI to use. Within these boundaries, disclosure, fidelity, accountability, and proportionality travel across modalities and settings; beyond them, the framework should be extended with care.

Faith-based communication as a stress test

Kenyan communicative life includes an unusually dense field of faith-based actors — churches, faith schools, and religious organisations; that are beginning to adopt the same generative tools under the same cost pressures as secular institutions, and the framework's four dimensions plausibly sharpen in that setting: disclosure carries the weight of honesty before a trusting community, cultural-linguistic fidelity extends to fidelity to a faith tradition and its language, human accountability becomes answerability for words and images carrying spiritual authority, and proportionality bears on how much of a sacred message may be delegated to a machine. The cases analysed here do not include a faith-based instance, and a dedicated extension of the framework to faith-based Kenyan communication, built on documented faith-based cases, is treated here as a distinct line of work rather than developed further in this paper.

Limitations and future directions

The study's limitations bound its claims. It is a conceptual paper that demonstrates rather than empirically tests its framework; its evidence is documentary and therefore confined to the public record; and its cases are drawn from a single national context, Kenya, treated as an information-rich entry point into dynamics shared across East Africa and the wider Global South. Future work should test and extend the framework through practitioner and audience data, across additional East African settings, and across generated text, audio, and multimodal outputs.

Conclusion

This paper set out to reconsider what authenticity requires of East African public relations practitioners in an environment increasingly mediated by generative AI, with particular attention to visual communication as the site where representational failures are most legible. Its central argument is that authenticity can no longer be treated as a property a message possesses or lacks, but must be understood as an ongoing practice, a set of deliberate ethical decisions taken each time a practitioner engages an AI tool. To make that practice concrete and teachable, the paper proposed a four-dimensional framework of algorithmic authenticity comprising disclosure, cultural-linguistic fidelity, human accountability, and proportionality, and demonstrated its diagnostic value through documented Kenyan cases of AI-generated visual content in institutional communication. Pursued further, algorithmic authenticity offers not a prohibition on generative tools but a disciplined way of using them: a reminder that, in the age of generative AI, authenticity is not something African institutions simply have or lose, but something their communicators must actively and repeatedly practise.

References

AIpots. (2024, January 18). Kenya Roads Authority using AI for their new adds annoys Kenyans. AIpots. https://www.aipots.com/trends/kenya-roads-authority-using-ai-for-their-new-adds-annoys-kenyans/

AlDahoul, N., Rahwan, T., & Zaki, Y. (2025). AI-generated faces influence gender stereotypes and racial homogenization. *Scientific Reports, 15*(1), 14449. https://doi.org/10.1038/s41598-025-99623-3

Alenichev, A., Kingori, P., & Peeters Grietens, K. (2023). Reflections before the storm: The AI reproduction of biased imagery in global health visuals. *The Lancet Global Health, 11*(10), e1496–e1498. https://doi.org/10.1016/S2214-109X(23)00329-7

Al Jazeera. (2026, July 31). *US government mislabels countries on map of Africa at global conference*. Al Jazeera. https://www.aljazeera.com/news/2026/7/31/us-government-mislabels-countries-on-map-of-africa-at-global-conference

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1493–1504). https://doi.org/10.1145/3593013.3594095

Birhane, A. (2020). Algorithmic colonization of Africa. SCRIPTed, 17(2), 389–409. https://doi.org/10.2966/scrip.170220.389

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Citizen Digital. (2024, January 18). Kenya Roads Authority used AI for their new ad and Kenyans were not impressed. Citizen Digital. https://www.citizen.digital/news/kenya-roads-authority-used-ai-for-their-new-ad-and-kenyans-were-not-impressed-n334949

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

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry, 12*(2), 219–245. https://doi.org/10.1177/1077800405284363

Gachie, K. (2024, January 31). Posta Kenya blasted online for sharing faulty AI image featuring extra hands and white employees. Citizen Digital. https://www.citizen.digital/article/posta-kenya-blasted-online-for-sharing-faulty-ai-image-featuring-extra-hands-and-white-employees-n335855

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review, 10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Kemibaro, M. (2024, January 12). Safaricom's AI-generated creative assets for advertising campaigns rile marketing practitioners in Kenya [Blog post]. Medium.

Kress, G., & van Leeuwen, T. (2021). Reading images: The grammar of visual design (3rd ed.). Routledge. https://doi.org/10.4324/9781003099857

Lim, J. S., & Jiang, H. (2022). Linking authenticity in CSR communication to organization–public relationship outcomes: Integrating theories of impression management and relationship management. *Journal of Public Relations Research, 33*(6), 464–486. https://doi.org/10.1080/1062726X.2022.2048953

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

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Moletsane, R. (2026). AI inclusivity and the centrality of Africa: A systematic review of representation, innovation, and governance in global artificial intelligence. *Frontiers in Artificial Intelligence, 9*, 1881299. https://doi.org/10.3389/frai.2026.1881299

Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report, 27*(1), 64–77. https://doi.org/10.46743/2160-3715/2022.5044

NTV Kenya. (2024). "Aiiii?" Posta Kenya reacts to AI advertisement blunder [Video]. YouTube. https://www.youtube.com/watch?v=JwlvIMlVP1o

Ooko, G. A. (2025). Toward a decolonial framework for communicating AI in African public service. Emerging Media, 3(3), 462–473. https://doi.org/10.1177/27523543251374174

Rocco, T. S., Plakhotnik, M. S., & Silberman, D. (2022). Differentiating between conceptual and theory articles: Focus, goals, and approaches. *Human Resource Development Review, 21*(1), 113–140. https://doi.org/10.1177/15344843211069795

Rwanda Inspirer. (2024). *Proposed media policy calls for labelling AI-generated content in Rwanda*. Rwanda Inspirer.

Sibbald, K. R., Phelan, S. K., Beagan, B. L., & Pride, T. M. (2025). Positioning positionality and reflecting on reflexivity: Moving from performance to practice. Qualitative Health Research. Advance online publication. https://doi.org/10.1177/10497323241309230

Stake, R. E. (1995). The art of case study research. Sage.

The Kenya Times. (2024, January 31). Posta forced to delete AI image after glaring mistake. The Kenya Times. https://thekenyatimes.com/latest-kenya-times-news/ai-gone-wrong-posta-forced-to-delete-ai-image-after-glaring-mistake/

TRT Afrika. (2026, August 2). *Mistake or intentional? US map of Africa blunder sparks outrage*. TRT Afrika. https://www.trtafrika.com/english/article/dd81cc3b0a11

Weinmann, H., Messingschlager, T. V., & Appel, M. (2026). Gender bias in text-to-image generative artificial intelligence: Neglect and stereotypical presentations across three popular platforms. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448261435197

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage.

Yang, Y. (2025). Racial bias in AI-generated images. *AI & Society, 40*, 5425–5437.