Journal of Visual and Performing Arts
Vol. 4 No. 1 (2026): ISSN (Online): 2957-5990
DOI: https://doi.org/10.58721/bdjz5t75

Whose Voice Speaks? Generative AI, Algorithmic Justice, and Strategic Communication in African Media

Cynthia Shimanyula
Kabarak University, Kenya

Article History
Received: 2026-02-20
Revised: 2026-08-17
Accepted: 2026-09-01
Published: 2026-09-08

Keywords
Africa
Algorithmic justice
Communication
Generative AI

How to cite:
Shimanyula, C. (2026). Whose Voice Speaks? Generative AI, Algorithmic Justice, and Strategic Communication in African Media. *Journal of Visual and Performing Arts*, 4(1), 58-69.

Abstract
Generative artificial intelligence (AI) is reshaping strategic communication across Africa, yet the tools are trained overwhelmingly on Global-North data and carry its representational defaults. This paper asks whose voice such systems amplify, and whose they silence, when African organisations use them to speak to their own publics. Using qualitative documentary analysis of publicly reported cases in which African organisations used generative AI in outward-facing communication and drew criticism on representational or authenticity grounds, and applying reflexive thematic analysis through a decolonial, algorithmic-justice lens, the study constructs three themes. First, a double bind of representation: uncritical AI fails along two opposite axes, erasing the local subject by rendering it as Global-North, or essentialising it by reducing it to stereotyped African signifiers, both rooted in training-data geography. Second, a political economy of authenticity, in which the turn to AI displaces local creative labour and audiences read the resulting communication as inauthentic partly because they infer that displacement. Third, the public as algorithmic watchdog, whereby affected publics, not regulators, detect and contest AI use, performing a grassroots decolonial critique. A contrast case of a brand publicly refusing AI underscores that authenticity has become a contested market value. The paper offers a conceptual model of algorithmic representational justice and argues that just AI use in African strategic communication requires re-centring the displaced local voice, not merely improving model outputs. Cases cluster in Kenya and South Africa, taken here as an information-rich entry point into a wider Global-South dynamic.

Introduction

Generative artificial intelligence (AI) has entered the everyday work of strategic communication across Africa, used by brands and public institutions to produce imagery, copy, and campaigns at speed and low cost. In the resource-constrained communication environments typical of the region, the appeal of tools that can generate a finished-looking advertisement from a text prompt is considerable, and adoption has been swift and largely ungoverned. But the leading systems that African communicators reach for are trained overwhelmingly on data drawn from the Global North, and they carry its demographic, aesthetic, and cultural defaults. When an African organisation uses such a tool to represent its own community, the question is not merely whether the output is accurate but whose assumptions it encodes, and at whose expense.

This paper asks a pointed version of that question: in AI-mediated African strategic communication, whose voice does the technology amplify, and whose does it silence? The question is deliberately one of power rather than only craft. A purely technical account would treat the misrendering of a Kenyan customer-service team, or the clichéd rendering of a South African audience, as quality-control problems to be fixed with better prompts or better models. A justice-oriented account, by contrast, asks why the defaults run in the direction they do, who benefits from the efficiency the tools promise, and who is written out of the communication and its production. It is the latter account this paper pursues.

The paper makes three contributions. Empirically, it assembles and analyses a corpus of documented cases in which African organisations used generative AI in their own outward-facing communication and drew public criticism on representational or authenticity grounds, a phenomenon that has been widely noticed in trade and popular commentary but not systematically analysed. Conceptually, it characterises the pattern in these cases as a coloniality of representation that operates along two opposite axes and proposes a model of algorithmic representational justice that links representation, labour, and public response. Normatively, it argues that the appropriate goal is not merely improved model outputs but the re-centring of the displaced local voice. It is positioned as an exploratory, theory-building study whose claims are bounded by a small, purposively assembled corpus and by a documentary method.

African organisations and practitioners are treated here as agents rather than as passive recipients of Global-North technology. In every case examined, the decision to deploy a generative-AI output in outward-facing communication was made by an identifiable Kenyan or South African team; in every failure case, the failure turned on a human step; cultural review, prompt revision, human-in-the-loop verification; that could have been taken and was not. Coloniality of the tool and practitioner responsibility are therefore held together in the argument that follows: both are true simultaneously, and neither excuses the other.

Literature Review
This study sits at the intersection of four literatures: the ethics and politics of algorithmic bias in generative systems; decolonial approaches to AI and technology; the theory of authenticity and trust in strategic communication and public relations (PR); and scholarship on media, representation, and the visual politics of Africa and the Global South.

Algorithmic bias in generative systems
A substantial body of empirical work now documents that generative AI systems reproduce and often amplify social biases present in their training data. Foundational audits of commercial vision systems established stark accuracy disparities across skin tone and gender (Buolamwini & Gebru, 2018), and subsequent syntheses confirm that AI systems trained on unrepresentative datasets produce discriminatory outcomes across race, ethnicity, language, gender, and geography (Ferrara, 2024). For generative text-to-image models specifically, controlled auditing has found that neutral occupational prompts return demographically skewed outputs and that these systems can amplify rather than merely mirror the skews in their training data (Sun et al., 2023).

More recent empirical work refines and, in important respects, complicates that early picture. Weinmann et al. (2026), in a preregistered content analysis of 1,344 images generated across DALL·E, Midjourney, and Stable Diffusion, document that both representational bias (whom the model chooses to depict) and presentational bias (how it depicts them) persist across the leading platforms, though the magnitude and shape of the bias vary considerably by prompt and by tool. AlDahoul et al. (2025), in a large-scale empirical audit published in Scientific Reports, extend the account to a phenomenon they term racial homogenisation, in which the model depicts members of a given race as strikingly similar to one another; crucially, the same study also demonstrates, in a preregistered experiment, that carefully designed debiasing interventions can materially reduce these effects. Yang (2025), in an empirical audit of image-to-image generation reported in AI & Society, extends the evidence beyond the text-to-image case: transforming 1,260 photographs of White, Black, and East Asian people, the study finds that White figures are reproduced with substantially higher racial accuracy than people of colour across single-person, same-race, and mixed-race conditions, with Black women depicted with the lowest accuracy in the single-person condition. The finding matters because it shows the erasure axis operating not only when models generate from a prompt but also when they transform an existing image of a subject unmistakably present in the input. A PRISMA-based systematic review by Moletsane (2026) consolidates the pattern across this literature, concluding that the limited representation of African populations, languages, and contexts in training data contributes directly to algorithmic bias. The direction of travel is twofold: bias in leading generative systems remains a documented and structural feature across both text-to-image and image-to-image pipelines, and yet targeted interventions, on the model and on the prompt, can measurably mitigate it. Any account that treats the technology as either irremediably biased or as already fixed misreads the evidence.

For the present argument, the most directly relevant evidence concerns the representation of Africa and the Global South specifically. In a widely-cited experiment, Alenichev et al. (2023) used a leading text-to-image tool to attempt to invert the entrenched "white saviour" tropes of global-health imagery, prompting for Black African doctors caring for white children. The tool proved structurally incapable of the inversion: it produced white doctors treating Black children even when prompted otherwise, and prompts such as "HIV patient" returned overwhelmingly Black subjects. Broader analyses reach compatible conclusions: text-to-image systems have been documented rendering African dwellings, even those of explicitly "wealthy" African subjects, as rural huts (Bianchi et al., 2023). A recent PRISMA-based systematic review consolidates this evidence, concluding that the limited representation of African populations, languages, and contexts in training data contributes directly to algorithmic bias (Moletsane, 2026).

Decolonial approaches to AI and technology

Where the bias literature documents what these systems do, decolonial scholarship theorises why it matters. Birhane (2020) argues that the large-scale deployment of Global-North-trained AI in Africa constitutes an algorithmic colonisation: externally built systems come to shape local social, cultural, and communicative life, while the data, the infrastructure, and the value accrue elsewhere. Birhane's (2021) relational-ethics account further argues that algorithmic injustice cannot be captured by decontextualised fairness metrics but must be understood in terms of concrete relationships, contexts, and harms.

Mohamed et al. (2020) provide the wider theoretical apparatus, showing how decolonial and postcolonial theory can function as a sociotechnical foresight tool. Applied to communication specifically, Ooko (2025) argues for a decolonial framework for communicating AI in African public service. Taken together, this literature offers the vocabulary, coloniality, extraction, re-centring, that the present study uses to interpret its cases.

Authenticity and visual representation

Authenticity in strategic communication is treated as foundational to the organisation–public relationship and to the trust on which reputation depends. Recent work demonstrates that perceived authenticity shapes relationship quality, credibility, and reputational outcomes, and that authenticity is not an intrinsic property of a message but a perception audiences form in context (Lim & Jiang, 2022). The study also draws on a long tradition of scholarship on how Africa and the Global South are represented in media and visual culture. That literature has documented how external image-making has repeatedly reduced the continent to a narrow repertoire of tropes—poverty, disease, wildlife, and an undifferentiated "African" aesthetic—and how such representation both reflects and reproduces relations of power. The significance of generative AI, from this vantage, is not that it invents new tropes but that it automates and scales old ones, embedding them in the routine production pipeline of advertising and institutional communication.

Across these literatures, the elements of an account exist but have not been assembled. The bias literature shows that generative systems misrepresent African subjects; decolonial theory explains why this is a matter of power; authenticity scholarship explains why publics reject it; and representation scholarship shows its historical continuity. What is missing is a study that brings these together to analyse the specific, documented practice of African organisations using these tools in their own strategic communication.

The gap this paper addresses

The elements of an account exist across the literatures reviewed above but have not been assembled for the specific case of AI-mediated African strategic communication. The bias literature shows that generative systems misrepresent African subjects; decolonial theory (Birhane, 2021; Mohamed et al., 2020) explains why this is not incidental but structural; the authenticity literature explains why misrepresentation is costly to organisations; and Ooko (2025) provides the closest available scaffolding for a decolonial framing of AI-related communication in African public service. Each of these leaves unfilled the specific space that a domain-specific analysis of AI-mediated African strategic communication would occupy: Birhane's relational ethics offers a general principle rather than a domain-specific diagnostic; Mohamed et al. offer a wide sociotechnical foresight lens rather than a model of practice; and Ooko's framework concerns communicating about AI rather than the ethics of using AI to communicate. The analysis that follows draws on those foundations to develop a domain-specific, prescriptive model of algorithmic representational justice for African strategic communication, one that connects representational, economic, and public-response dimensions into a single diagnostic-and-corrective account.

Method

This section sets out the study's design, corpus, and analytic procedure. The exposition proceeds through subsections that follow the standard order: paradigm and design, corpus construction and inclusion criteria, evidential grading, analytic approach, and reflexivity, ethics, and limitations.

Design and paradigm

The study uses qualitative documentary analysis, an established method in which publicly available texts and artifacts are examined systematically to generate interpretive insight (Bowen, 2009; Morgan, 2022). It is positioned within an interpretivist paradigm with a critical, decolonial orientation: representation is treated as a site of power, and the analysis reads AI-generated artifacts and their public reception as evidence of how that power operates. The design is exploratory and theory-building; it constructs themes and a conceptual contribution rather than testing hypotheses or claiming statistical generalisation.

The cases are engaged through secondary sources: published news reports, trade and practitioner commentary, and documented public reactions on visible platforms. There is no primary access to the organisations concerned, to the internal decisions behind their 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. Qualitative documentary analysis is designed for exactly this kind of reading, and the corpus and analytic procedure that follow are structured around what public documentation can and cannot support.

Corpus construction and inclusion criteria

Cases were assembled through systematic searching of public sources, including national and pan-African news reporting, trade and practitioner commentary, and fact-checking and civil-society documentation, and screened against explicit criteria. A case was included only where an identifiable African organisation, brand, or institution deliberately used generative AI in its own outward-facing strategic or brand communication, and the output drew documented public criticism on representational, cultural, linguistic, or authenticity grounds, corroborated by at least one credible, datable source. Searches were iterative and continued until additional searches within the defined scope ceased to surface qualifying new cases.

The corpus deliberately excludes malicious deepfakes, impersonation, and AI-enabled scams, in which bad actors fabricate content featuring a person or brand without their involvement. Although abundant across the continent, this category belongs to a different phenomenon, disinformation and fraud, rather than the ethics of an organisation's authentic self-representation. Drawing this boundary is itself a methodological contribution, since much popular discussion of "AI in African media" runs the two together.

A notable feature of corpus construction is itself a finding worth reporting. Searches across Ghana, Uganda, Tanzania, and Nigeria returned almost exclusively the excluded deepfake category, while cases of organisational AI self-representation clustered in Kenya and South Africa, the continent's deepest digital-advertising ecosystems and home to some of its most vocal online publics. The absence of Ghanaian, Ugandan, Tanzanian, and Nigerian cases from the analytic corpus does not license the inference that organisations in those countries do not use generative AI in outward-facing communication, nor that when they do the use is culturally consonant and therefore uncontroversial. Trade reporting and practitioner commentary from all four countries indicate that generative AI is being adopted for brand and institutional communication in those markets, and there is no evidence on which to characterise that adoption as uniformly culturally consonant. What the present searches suggest instead is that in those countries such use, when it occurs, either does not yet attract the volume of publicly documented critical response that would surface cases in a search of the kind conducted here, or attracts responses that circulate on platforms and in languages that an English-language, mainstream-media-weighted search does not adequately reach. The pattern is thus an artefact of the visibility of critical public response, mediated by digital-advertising market depth and by the size and volubility of the critical online public sphere, rather than a claim about the underlying rate of AI use. Kenya and South Africa accordingly serve as an information-rich entry point into a wider Global-South dynamic, a scoping decision grounded in what could be documented rather than in a substantive claim about the excluded countries.

Evidential grading and the final corpus

Because documentary sources vary in credibility, cases were graded by evidential strength. Four cases carry the primary analytic weight. Three are Kenyan cases of representational failure: the Postal Corporation of Kenya (Posta Kenya), whose AI-generated customer-care image rendered part of a Kenyan state corporation's own staff as white and displayed anatomical anomalies before being withdrawn without public explanation; the Kenya Urban Roads Authority (KURA), whose AI-generated road-connectivity imagery was judged culturally implausible and mocked by audiences; and Safaricom, whose broad shift of creative production to generative AI drew alarm from the Kenyan marketing community. The fourth is a South African contrasting case, Windhoek Beer's public pledge to refuse AI-generated imagery, which runs in the opposite direction and clarifies the stakes.

A further set of corroborating cases, documented by single sources or as reported trends, is used illustratively rather than as load-bearing. These include a documented shift toward AI creative on Nairobi billboards and a cluster of South African brands criticised in trade commentary for AI advertising built on stereotyped "African" tropes.

Analytic approach

The corpus was analysed using reflexive thematic analysis (Braun & Clarke, 2006, 2021), an interpretive approach in which patterns of meaning are actively constructed by the researcher through systematic, recursive engagement with the material. Analysis followed the familiar recursive phases: familiarisation with each case and its documented reception; initial coding of representational features, stated rationales, and public responses; the construction of candidate themes; and their review, refinement, and definition. The reading was theory-informed but open: the decolonial and algorithmic-justice literature sensitised the analysis, while the themes were developed from the cases rather than imposed as a predetermined coding frame.

Reflexivity, ethics, and limitations

The author writes as a Kenyan PR practitioner and communication educator. This standpoint is treated as an analytic resource: it affords contextual familiarity with the registers, cues, and expectations against which the AI-generated artifacts are judged culturally implausible. It also carries interpretive risk, mitigated by grounding every claim in the corroborated public record, by grading cases evidentially, and by stating limits openly.

Ethically, the study analyses publicly available organisational communications and publicly visible reactions to them; it involves no human participants, no private data, and no interaction with individuals. Its limitations follow from its design: as documentary analysis, it reads artifacts and public traces and cannot access organisations' internal decisions; and its corpus is small, purposively assembled, and weighted toward Kenya and South Africa

Results

Three themes were constructed from the corpus, together with a contrast case and a synthesising observation that answers the paper's guiding question. Each theme is presented with the cases that ground it and with attention to the evidential strength of those cases.

The anchor cases

The first case, Posta Kenya, involved the Postal Corporation of Kenya publishing on the platform X a promotional image for its customer-care service: four smiling agents in headsets. Kenyan users rapidly identified that the image was AI-generated, noting anatomical anomalies (figures with supernumerary hands) and, more consequentially, that part of the depicted staff of a Kenyan state corporation had been rendered as white. The corporation withdrew the image after a short interval and offered no public account of what had occurred (Gachie, 2024). The case is analytically rich because a single artifact exhibits erasure (the misracing of local staff), an accountability vacuum, and the public performing the detection that the institution did not.

The second case, the Kenya Urban Roads Authority (KURA), involved a statutory public body publishing AI-generated images promoting national road connectivity. Audiences judged the imagery culturally and materially implausible—roads terminating without destination, scenes stripped of the amenities of an actual Kenyan streetscape—and mocked both the imagery and the decision to use AI for official public communication (Citizen Digital, 2024; AIpots, 2024). The third case, Safaricom, is not a single artifact but a documented shift: the country's most prominent brand moved much of its creative production to generative AI, prompting alarm across the Kenyan marketing community (Kemibaro, 2024).

The fourth anchor case, Windhoek Beer, is the contrast: a South African brand that publicly pledged not to use AI-generated imagery and built a campaign around the authentically human. Together, the four cases span the relevant range: a corporate and a governmental instance of representational failure, a brand-level account of labour displacement, and a counter-instance of deliberate refusal.

Theme one: The double bind of representation—erasure and essentialism

The corpus reveals two opposite modes of representational failure, both traceable to the same root, and their co-presence is the study's central analytic finding. In the first mode, erasure, generative tools default to Global-North demographics and aesthetics and write the local subject out of the frame. The clearest documented instance is the Posta Kenya image, in which part of a Kenyan state corporation's own customer-care staff was rendered as white (Gachie, 2024; Citizen Digital, 2024).

In the second mode, essentialism, the tools render "Africa" not by erasing it but by reducing it to a narrow repertoire of stereotyped signifiers. South African commentary describes brands prompting for "Africanacity" clichés—beads, braids, dreadlocks, generic "African" dress, and safari scenes—producing a flattened, touristic version of local identity (GLITCHED, 2025; Businessday, 2025). This is the essentialising counterpart to erasure, and it is corroborated by peer-reviewed evidence that generative systems default to stereotyped depictions of non-Western subjects (Bianchi et al., 2023).

These are opposite surface failures—one absents the local subject, the other over-marks it—yet they share a single cause: a model trained predominantly on Global-North data can neither represent the local subject faithfully nor imagine it beyond cliché. The peer-reviewed evidence reinforces the point at a structural level: Alenichev et al. (2023) found a leading image tool unable to depict Black African professionals caring for white patients even when explicitly prompted. What the African communication cases show in the wild, the audit literature shows under controlled conditions: uncritical generative AI fails on the African subject along two structured, opposite axes.

Theme two: Authenticity and the political economy of creative labour

Across the corpus, the turn to generative AI is repeatedly cost-driven, and the loss of authenticity is entangled with the displacement of local creative labour. The Safaricom case was framed by a Kenyan practitioner precisely as a substitution of AI for the models, actors, photographers, and art directors whose work had made the brand's communication recognisably Kenyan; the same practitioner later extended the observation to a broader shift toward AI creative on Nairobi billboards, asking where the "real humans" had gone (Kemibaro, 2024, 2025). Kenyan commentary framed the KURA and Safaricom episodes explicitly as a tension between technological innovation and local employment (Mutistarink, 2024).

Authenticity here is not only a property of the image, whether it "looks real," but a trace of the human and economic relations behind it. Audiences read AI-generated brand communication as inauthentic partly because they infer, correctly, that local people were written out of its production, and that the value once distributed to local creative workers has been captured elsewhere. This connects the representational coloniality of Theme One to the extractive political economy that decolonial scholarship places at the centre of algorithmic power (Birhane, 2020): the same move that erases or caricatures the local subject in the image also removes the local worker from the pipeline that would have produced it.

Theme three: The public as algorithmic watchdog

In every firmly-documented case, it was the public, not a regulator or the organisation itself, that detected the AI, named its failure, and imposed a reputational cost. Kenyans on the platform X identified the anomalies in Posta's image and mocked the implausibility of KURA's roads (Gachie, 2024; Citizen Digital, 2024); South African audiences flooded brand AI advertising with criticism and coined a vernacular of "AI slop" and "sell-outs" (GLITCHED, 2025). Strikingly, this public critique is frequently decolonial in substance if not in name. To observe that even KURA's AI "fantasies" cannot picture a properly serviced Kenyan road, or to reject the "Africanacity" cliché as a lazy externalised idea of Africa, is to perform from below the same critique of algorithmic coloniality that the scholarly literature makes from above.

This finding carries both hope and caution. It is hopeful because it shows agency that decolonial scholarship, focused on the power of systems, sometimes underplays: publics can and do impose real costs on organisations that misrepresent them. It is cautionary because reliance on public backlash as the de facto accountability mechanism is unreliable and unevenly distributed. It protects best precisely those audiences already most AI-literate, most connected, and most vocal, and fails silently where those conditions are absent. The pattern therefore strengthens the case for institutional disclosure norms and for teachable practitioner frameworks that do not depend on after-the-fact public outrage.

A contrast case: refusing AI as brand value

One case in the corpus runs against the others and clarifies what is at stake. Windhoek Beer, a South African brand, publicly formalised a "100% Real" pledge never to use AI-generated imagery in consumer-facing communication, and built a campaign around a Kenyan content creator with polydactylism—six fingers on each hand—precisely to weaponise the notorious AI "extra fingers" giveaway as a mark of the authentically human (Windhoek Beer, 2026).

The case is analytically valuable for two reasons. First, it demonstrates that authenticity in AI-mediated communication has become a contested market value, one real enough that a brand can differentiate itself by publicly refusing the very technology its competitors are adopting. Second, its choice of a Kenyan human face to stand for the "real" ties the corpus's two national strands together and enacts the alternative the failure cases lack: where Posta's AI erased the Kenyan worker and the "Africanacity" ads caricatured the African subject, Windhoek's campaign re-centres a specific, real, local person, and makes his very particularity, his six-fingered hands, the point.

Positive cases: proper use of AI in African communication

The failure cases and the refusal case sit within a wider field of African AI-communication practice that includes documented examples of considered, culturally attentive use. The Mbaza Chatbot, deployed in Rwanda during the COVID-19 response, was a locally-developed AI communication tool designed around the linguistic, informational, and trust needs of the Rwandan public (Ooko, 2025). The Masakhane NLP project, an African-led open science initiative building natural language processing resources for African languages, works at the infrastructural level: rather than continuing to receive Global-North models trained without African languages, African researchers and practitioners are building that linguistic capacity themselves, addressing at source the training-data problem that produces the representational failures analysed above. Read alongside the failure cases, these examples sharpen rather than dilute the pattern: generative AI in African strategic communication is unworkable where practitioner responsibility, cultural review, and community participation are absent, and workable, sometimes valuable, where they are present. Human judgment and local participation are what distinguish the two.

Synthesis: whose voice speaks?

Read together, the three themes and the contrast case answer the paper's title question. In uncritical AI-mediated African strategic communication, three voices are in play, and the injustice lies in their relation. The model's voice, shaped by Global-North data, speaks loudest, supplying the defaults that erase or caricature the local subject. The local creative's voice is displaced, written out of production by cost-driven substitution. And the public's voice rises in protest, performing the detection and critique that institutions and regulators have not. The contrast case shows that the alternative, letting the real local voice speak, is available and even advantageous. The paper's contribution is to name this configuration and to reframe the objective: algorithmic justice in African strategic communication is not chiefly a matter of debiasing model outputs, but of re-centring whose voice gets to speak, and whose labour gets to make, a community's public image.

A conceptual model of algorithmic representational justice

The findings can be drawn together into a conceptual model that links the three themes and specifies the relations among them. The model has three interacting components and a corrective vector. The first is the representational mechanism: training-data geography produces a double bind in which the local subject is either erased or essentialised. The second is the economic mechanism: cost-driven adoption displaces local creative labour, removing the community's participation in the production of its own image. These two mechanisms are mutually reinforcing, and together they constitute what the paper terms algorithmic representational justice. The third component is the corrective response: affected publics detect and contest the injustice, functioning as an emergent, uneven accountability mechanism. The corrective vector, indicated by the contrast case, is re-centring: the deliberate return of the real local subject and the local creative to, respectively, the image and its production.

The model's value is diagnostic and prescriptive. Diagnostically, it directs analysts and practitioners to look for both failure axes and for the labour displacement behind them. Prescriptively, it implies that adequate responses cannot be confined to output-auditing or disclosure but must also address the distributive question of who makes the communication.

Discussion

These findings extend algorithmic-justice scholarship from its usual domains—search, recognition, moderation—into strategic communication and branding, showing that the coloniality of representation operates not only in consequential state and platform systems but in the everyday commercial image-making through which publics most often meet mediated versions of themselves. The double-bind finding, in particular, refines the standard account: it shows that algorithmic bias against African subjects is not a single undifferentiated failure but a structured pair of opposite failures, and that a theory or an audit attending to only one of them will miss half of the phenomenon.

The political-economy finding complicates the remedies most often proposed. Disclosure norms and cultural quality-assurance address the representational harm but not the distributive one. If audiences reject AI communication partly because it displaces local creative labour, then a fully-disclosed, culturally-accurate AI image may still be read as inauthentic, because the objection was never only to the image. This aligns the communication debate with the broader decolonial critique of AI as an extractive formation.

The watchdog finding speaks to the governance of AI in thin-guardrail environments. Where formal regulation of AI in advertising is nascent or absent, the affected public has become the primary enforcement mechanism. This is a fragile foundation for accountability: it is reactive, uneven, and dependent on the very digital literacy and connectivity that are unequally distributed. The finding therefore supports investment in more durable arrangements, professional disclosure standards, and curricula that equip the next generation of communicators to audit AI outputs. At the same time, the reality of an active, critical public is a resource that such arrangements can build upon rather than replace.

Situating the findings within the older literature on the representation of Africa underscores their significance. The tropes that generative AI reproduces—the erased professional, the essentialised "African," the suffering Black subject—are not new; they are the automation, at scale and speed, of a visual politics that critical media scholarship has contested for decades. What is new is the pipeline: these representations now enter public circulation not through the deliberate choices of external image-makers but through the routine, prompt-and-generate workflows of local organisations using global tools.

The human factor: from tool critique to practice critique

Where responsibility for the documented failures lies is analytically distinct from what the tools themselves make possible or difficult, and the distinction matters for what follows. The failures analysed above are not evidence that the tools are inherently unusable in African contexts; they are evidence that the tools are unusable in the way they were used — without cultural review, without prompt revision, without human-in-the-loop verification, and without evident audit of the outputs against the audiences those outputs would reach. Every failure in the corpus turned on a human decision, at a Kenyan or South African desk, to skip a step of practitioner responsibility that would have caught the problem. Two claims must therefore be held separately. The first is that leading generative systems carry biases inherited from their training data; the recent audits summarised above support this claim with strong empirical evidence. The second is that African organisations are helpless in the face of that bias. The cases here weigh against the second claim: the corrective disciplines developed later in the argument — cultural fidelity, participatory production, institutional accountability — are ordinary practitioner competences that were within reach in each case and were not exercised. Both claims are true simultaneously; neither excuses the other.

Implications for practice, pedagogy, and policy

For practitioners, the double-bind finding translates into a concrete auditing discipline: outputs must be checked not only for the obvious failure of erasure but also for the subtler failure of essentialism, in which the tool produces a technically "African" image that is nonetheless a stereotype. A cultural quality-assurance step that screens for both, ideally involving members of the represented community, is a minimal safeguard. For educators, teaching the next generation of communicators to use AI responsibly cannot be reduced to prompt-craft; it must include the critical capacity to recognise how these tools encode representational politics. For policy and industry bodies, the watchdog finding argues for durable accountability arrangements that do not leave redress to the uneven mechanism of public backlash.

Limitations

The corpus is small, purposively assembled, and concentrated in two countries; while that concentration is itself an evidenced pattern, it means the findings speak most confidently to Kenya and South Africa. Several corroborating cases rest on single, sometimes non-scholarly sources and are used illustratively; the load-bearing claims rest on the firmly-documented anchor cases and the peer-reviewed audit literature. As documentary analysis, the study cannot access the internal reasoning of the organisations concerned. And as an interpretive, reflexive-thematic study authored by a practitioner-scholar with a stake in the field, it makes no claim to a view from nowhere; its account is defensible, transparent, and grounded, but it is an interpretation.

Directions for future research

The double-bind and watchdog findings could be tested against a larger, more systematically assembled, and more geographically distributed corpus. The representational claims could be triangulated with a controlled outputs audit, generating images from African communication prompts across current tools and coding them for erasure and essentialism. The political-economy claim could be pursued through interviews with African creative workers and agencies. And the watchdog and contrast findings could be examined through audience research, testing whether the authenticity premium that made refusal viable for one brand generalises.

Conclusion

Asked whose voice speaks in AI-mediated African strategic communication, this study answers that, used uncritically, generative AI amplifies a Global-North-shaped model voice, displaces the local creative voice, and provokes a public voice of protest. Through documentary analysis of Kenyan and South African cases, read with reflexive thematic analysis and a decolonial lens, it characterises this as a coloniality of representation operating along two opposite axes—erasure and essentialism—entangled with the displacement of local creative labour and contested from below by publics acting as algorithmic watchdogs.

Drawing these together, the paper proposes a model of algorithmic representational justice and argues that the appropriate goal is not merely to debias model outputs but to re-centre the displaced local voice, in the image and in its making. That one brand could build authenticity by publicly refusing AI, and could stake its "real" on the particular, six-fingered hands of a specific local person, suggests that the displaced voice can be re-centred, and that in a market where authenticity has become contested, communicators may find that letting it speak is not only just but wise.

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

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

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

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

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

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html

Businessday. (2025, September 9). How AI will affect South Africa's advertising and communications industry. Business Day. https://www.businessday.co.za/redzone/news-insights/2025-09-09-how-ai-will-affect-south-africas-advertising-and-communications-industry/

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

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), 3. https://doi.org/10.3390/sci6010003

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

GLITCHED. (2025). South Africa has an AI ad pandemic and I am so happy everyone else hates it too. GLITCHED.

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

Kemibaro, M. (2025, December 19). Where have all the real humans gone from Kenya's billboards? [Blog post]. https://moseskemibaro.com/2025/12/19/where-have-all-the-real-humans-gone-from-kenyas-billboards/

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.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684.

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

Mutistarink. (2024). AI in Kenyan advertising: Navigating change and controversy. https://www.mutistarink.com/ai-journal/ai-in-kenyan-advertising-impact-and-debate

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

Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2023). Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI https://doi.org/10.1093/jcmc/zmad045

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

Windhoek Beer. (2026, June). Windhoek Beer commits to human-generated imagery in new marketing campaign ["The Real Guy" campaign]. Ad Age. https://adage.com/creativity/work/aa-extra-fingers-anti-ai-windhoek-beer/

Yang, Y. (2025). Racial bias in AI-generated images. *AI & Society, 40*, 5425–5437. https://doi.org/10.1007/s00146-025-02282-1