



AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space

Okpaka Chukwuma FNCS

Department of Computer and Information Technology, Veritas University, Abuja, Nigeria.

OPEN ACCESS

Article Citation:

Okpaka Chukwuma FNCS, "AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space", International Journal of Recent Trends in Multidisciplinary Research, July-August 2026, Vol 6(04), 385-395.



©2026 The Author(s). This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and

reproduction in any medium, provided the original author and source are credited. Published by 5th Dimension Research Publication

Abstract: Generative artificial intelligence (AI) has introduced a new and structurally distinct class of threat to electoral integrity in Nigeria: the capacity to produce hyper-realistic synthetic media — including voice-cloned audio, deepfake video, and AI-generated text — in Nigeria's three dominant regional languages (Hausa, Yoruba, and Igbo) and in Nigerian Pidgin, at negligible cost and without technical expertise. Nigeria's 2023 general election constituted a watershed moment in this regard, marked by a documented proliferation of deepfake content falsely implicating presidential candidates in ballot manipulation, ethno-religious sedition, and criminal conduct. As Nigeria approaches the 2027 general elections, the threat has escalated in both sophistication and scale. This paper investigates the mechanisms, vectors, and political economy of AI-mediated electoral manipulation in Nigeria's multi-ethnic digital space, with particular attention to the intersection of generative AI and ethno-regional political identity. Drawing on a mixed-methods study comprising content analysis of 1,240 synthetic media artefacts identified during the 2023 election cycle, semi-structured interviews (n=48) with fact-checkers, platform moderators, civil society actors, and INEC officials, and a structured digital survey (n=720) administered across Nigeria's six geopolitical zones, the paper develops a typology of AI-mediated electoral manipulation (AMEM) and proposes a four-pillar Democratic Consent Protection Framework (DCPF) for Nigeria. Findings reveal that 67.4% of surveyed registered voters encountered AI-generated political content in the 30 days preceding the 2023 election, that ethnic and religious targeting was the dominant manipulation modality (present in 71.3% of identified deepfake content), and that existing regulatory instruments — including the Electoral Act (2022), the Nigeria Data Protection Act (2023), and INEC's nascent AI Division — are structurally inadequate to address the speed, scale, and linguistic specificity of AI-mediated electoral threats. The paper contributes the first empirically grounded typology of AI electoral manipulation in a Sub-Saharan African multi-ethnic democracy, and the first governance framework specifically designed for this context.

Keywords: AI Electoral Manipulation; Deepfakes; Nigeria; Democratic Consent; Multi-Ethnic Democracy; Disinformation; Generative AI; 2027 Elections; DCPF Framework

1. Introduction

The integrity of democratic consent, aptly described as the foundational principle that electoral outcomes reflect the

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space

uncoerced and informed preferences of citizens — is undergoing an unprecedented technological stress test. In Nigeria, the emergence of generative AI as a low-cost, high-fidelity instrument of political manipulation has introduced a structurally new threat to democratic legitimacy that existing legal, institutional, and civil society frameworks are not equipped to address. This paper investigates that threat in the specific context of Nigeria's multi-ethnic digital space, where the intersection of over 250 ethnic groups, three regionally dominant languages, a history of ethno-religious electoral violence, and rapidly expanding social media penetration creates a manipulation environment of exceptional complexity and fragility.

Nigeria's 2023 presidential election was, by multiple scholarly and journalistic accounts, the world's first national election in which AI-generated deepfake content played a documented and significant role in the information environment [1, 2]. An AI-manipulated audio clip falsely implicated a presidential candidate in plans to manipulate ballots; deepfake videos circulated depicting candidates in ethno-religious contexts designed to inflame Hausa, Yoruba, and Igbo voter blocs; and AI-generated text content was deployed via WhatsApp voice notes — the dominant political communication medium in Nigeria — to amplify false narratives at speed and scale impossible to human-generate manually [3, 4]. These incidents were not isolated; a structured content analysis of disinformation reported by Nigerian fact-checking organisations during the election cycle identifies over 1,200 distinct AI-generated or AI-amplified content artefacts [5].

The threat is escalating. As Nigeria approaches its 2027 general elections, AI systems capable of generating convincing synthetic speech in Hausa, Yoruba, Igbo, and Nigerian Pidgin have become widely accessible; the technical barrier to producing a convincing deepfake of a political figure addressing their ethnic constituency in their mother tongue has collapsed [6, 7]. A coordinated AI swarm operation that disseminates ethnically tailored disinformation simultaneously across Nigeria's six geopolitical zones from a single command node would have required enormous human resources five years ago; today it is within the operational capacity of a small, modestly resourced political actor [8].

Nigeria's regulatory architecture has not kept pace with this evolution. The Electoral Act (2022), the Nigeria Data Protection Act (2023), the Cybercrimes (Prohibition, Prevention, Etc.) Act (2015), and the NITDA AI Policy (2024) collectively address elements of the digital information environment, but none establishes a framework adequate to the speed, scale, linguistic specificity, and multi-platform character of AI-mediated electoral manipulation [9, 10]. In May 2025, the Independent National Electoral Commission (INEC) established an Artificial Intelligence Division mandated to combat electoral disinformation; experts have welcomed this step while cautioning that institutional capacity, funding, and detection technology remain insufficient [11].

Against this backdrop, this paper pursues four objectives: (i) to construct the first empirical typology of AI-mediated electoral manipulation (AMEM) in Nigeria; (ii) to document the ethnic and regional targeting patterns of AI-generated electoral disinformation during the 2023 election cycle; (iii) to assess the adequacy of Nigeria's existing regulatory architecture for addressing AI electoral manipulation; and (iv) to propose and validate a four-pillar Democratic Consent Protection Framework (DCPF) for Nigeria. The paper's empirical contributions draw on content analysis of 1,240 media artefacts, 48 semi-structured interviews, and a structured survey of 720 registered voters across Nigeria's six geopolitical zones.

The paper proceeds as follows. Section 2 reviews the literature on AI and electoral integrity, democratic consent theory, and disinformation in multi-ethnic contexts. Section 3 details the research methodology. Section 4 presents findings on the typology and targeting patterns of AI electoral manipulation in Nigeria. Section 5 analyses regulatory adequacy. Section 6 develops and validates the DCPF. Section 7 discusses implications and limitations. Section 8 concludes.

2. Literature Review

2.1 Democratic Consent and Its Digital Vulnerabilities

Democratic consent, in the Rousseauian tradition elaborated through Rawls [12] and Habermas [13], rests on the premise that legitimate democratic outcomes require conditions of free, informed, and undistorted political communication among citizens. Habermas's concept of the ideal speech situation is that in which all participants have equal communicative access and all claims are subject to rational scrutiny. This concept provides the normative benchmark against which AI-mediated manipulation can be evaluated: synthetic media, bot-amplified disinformation, and algorithmic micro-targeting each corrupt the communicative conditions that democratic consent requires. Zuboff [14] extends this analysis to the contemporary digital context, demonstrating how the infrastructure of digital communication has become an instrument of behavioural manipulation at scale, with profound implications for democratic self-determination.

In ethnically plural societies, the vulnerability of democratic consent to information manipulation has additional dimensions. Horowitz [15] demonstrated that ethnic identity is a primary organising principle of political competition in deeply divided societies, and that political actors have systematic incentives to mobilise ethnicity as a voting cue. Brass [16] developed the concept of 'instrumental ethnicity,' arguing that political elites strategically construct and deploy ethnic narratives to consolidate electoral support. AI-generated content that reinforces or fabricates ethnic narratives can therefore function not merely as disinformation in the generic sense, but as a targeted instrument of ethno-political manipulation that bypasses rational deliberation and activates identity-based political reflexes.

2.2 AI and Electoral Integrity: Global Evidence

The global evidence on AI's impact on electoral integrity is rapidly accumulating. During Slovakia's 2023 parliamentary elections, AI-generated audio recordings impersonated a liberal candidate discussing plans to raise alcohol prices and rig the election, circulating on Facebook days before the poll [17]. In Bangladesh's January 2024 elections, a deepfake video depicting a candidate announcing her withdrawal circulated hours before polls opened [18]. In Indonesia's 2024 presidential race,

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space

deepfake videos falsely showed a candidate speaking Arabic to appeal to Muslim voters [19]. The February 2025 German federal election witnessed the 'Storm-1516' operation, which deployed networks of AI-generated websites to distribute deepfake content attacking candidates [20]. In each case, the defining characteristic was speed: by the time content was debunked, the damage to voter perception had already occurred.

The scholarly literature has begun to theorise these incidents. Vaccari and Chadwick [21] introduce the concept of 'deepfake uncertainty,' demonstrating that synthetic media does not need to be widely believed to be politically effective; it suffices for it to generate epistemic confusion that makes voters distrust all political information, including accurate information. Wardle and Derakhshan [22] distinguish between misinformation (false content shared without harmful intent), disinformation (false content shared with harmful intent), and malinformation (true content shared to cause harm), arguing that AI-generated electoral content primarily constitutes disinformation. Bradshaw and Howard [23] document the industrialisation of computational propaganda, described as the use of automated accounts and algorithmic amplification to manufacture artificial social consensus, as a global electoral threat.

2.3 AI Electoral Manipulation in Nigeria: The 2023 Precedent

Nigeria's 2023 presidential election has generated a small but significant body of documentation that establishes it as a global precedent in AI-mediated electoral manipulation. Ajakaiye [1], in testimony reported by the Institute for Security and Technology, described the scope of AI-generated disinformation during the 2023 cycle as 'beyond the resources of local journalists to address,' and argued that 'the Nigeria presidential election was perhaps the world's first instance of a country's democratic process being marred by such means.' The Africa in Fact analysis [3] documented specific deepfake deployments: AI-generated images and videos falsely linked candidates to terrorist groups and 'ethnocentric dialogues'; deepfake videos depicted a candidate incoherently responding to questions at Chatham House; and AI voice-synthesised content purporting to show candidates endorsing ballot manipulation circulated widely on WhatsApp.

The ethnic targeting dimension of this manipulation is documented by Dubawa [4], which identified a coordinated ethno-disinformation network that 'recycles old news, distorts media reports, and fabricates claims to deepen ethnic divisions' and which 'often portrays the Igbos, Yorubas, and Hausa-Fulanis as inherently violent, criminal, or unfit for political leadership.' This network was not exclusively AI-generated, but AI tools significantly amplified its reach and lowered the production cost of its content.

The linguistic dimension amplifies these risks. The Cable [7] reports that political communication in Nigeria is increasingly conducted through WhatsApp voice notes in local languages, and that 'AI technology can automatically translate and synthesise speech in several languages, including local languages, which enables the creation of convincing political communication that sounds like respected figures in the local community, religious leaders, traditional rulers, or local politicians.' Legit.ng [8] notes that a coordinated AI disinformation operation 'could disseminate regionally tailored disinformation simultaneously, with Hausa language content stoking tension in Kaduna while Yoruba language content fuels unrest in Ibadan, all from a single coordinated command.'

2.4 Regulatory Frameworks and Their Gaps

Existing scholarship on AI governance in Nigeria and Africa has documented the general inadequacy of current regulatory frameworks for responsible AI deployment [24, 25]. In the electoral context, this inadequacy is compounded by the speed at which AI-generated content must be identified, assessed, and removed to prevent electoral harm. The Electoral Act (2022) addresses traditional electoral offences including bribery, intimidation, and impersonation, but contains no provisions adequate to synthetic media, algorithmic micro-targeting, or bot-driven disinformation [9]. The Cybercrimes Act (2015) criminalises the use of computers for fraud and harassment but predates generative AI and contains no provisions for synthetic political media [10]. Nigeria's nascent AI Policy (2024) addresses responsible AI deployment in general terms but does not establish an electoral-specific AI governance regime [26].

Comparatively, Techpoint Africa [6] documents that 'Nigeria currently lacks a national AI governance framework, a dedicated election cybersecurity unit, or a coordinated rapid response system for electoral disinformation,' while NITDA and the National Cybersecurity Coordination Centre 'will need to significantly expand their mandates and capabilities before 2027.' BusinessDay [11] notes that 'Nigeria's regulatory framework has not yet fully adapted to the realities of AI. While existing laws address traditional misinformation, they do not adequately cover the complexities of synthetic media and algorithmic manipulation.' This paper builds on these observations by providing the first systematic regulatory gap analysis specific to AI electoral manipulation in Nigeria.

3. Materials and Methods

3.1 Research Design

This study employs a sequential mixed-methods design integrating three complementary strands of evidence: systematic content analysis of AI-generated electoral disinformation artefacts, qualitative semi-structured interviews with expert and institutional informants, and a structured quantitative survey of registered voters. The design is grounded in a constructivist epistemological position that recognises AI-mediated electoral manipulation as a socially constructed threat whose meanings and effects are produced through the intersection of technology, political identity, and institutional context [27]. Quantitative data are used to characterise the scale and distribution of the phenomenon; qualitative data illuminate its mechanisms and meanings.

3.2 Study Design

Sequential exploratory mixed-methods design, comprising three phases: Phase 1 (content analysis); Phase 2 (qualitative interviews); Phase 3 (voter survey and framework development).

3.3 Study Location

The study was conducted across Nigeria's six geopolitical zones: North-West (Kano), North-East (Maiduguri), North-Central (Abuja/Nasarawa), South-West (Lagos and Ibadan), South-East (Enugu and Onitsha), and South-South (Port Harcourt and Benin City). This geographic spread was designed to capture Nigeria's full ethnic, linguistic, and political diversity, and to enable zone-by-zone comparison of AI manipulation exposure and impact.

3.4 Study Duration

Phase 1 content analysis: Months 1–8. Phase 2 qualitative interviews: Months 6–12 (concurrent with Phase 1 coding to enable iterative refinement). Phase 3 voter survey and framework validation: Months 13–22. Total study duration: 22 months (November 2023 – August 2025).

3.5 Sample Size

Phase 1 content analysis corpus: $n=1,240$ synthetic media artefacts collected from fact-checking databases of Dubawa, AFP Fact Check Nigeria, Peoples' Gazette, and the BBC Africa Eye archive for the period October 2022 – March 2023. Phase 2 qualitative interviews: $n=48$ participants. Phase 3 voter survey: $n=720$ registered voters across six geopolitical zones (120 per zone). Delphi validation panel: $n=22$ experts.

3.6 Sample Size Calculation

Phase 3 survey sample size was calculated using the single proportion formula at 95% confidence interval and 5% margin of error, with an assumed population proportion of 0.5 (maximum variance). The registered voter population was taken as 93.5 million (INEC, 2023). The minimum required sample was $n=384$; 720 was targeted to enable meaningful zone-level disaggregation ($n=120$ per zone) and to accommodate an anticipated 12% non-response rate.

3.7 Subjects and Selection Method

Phase 1 artefacts were identified through systematic keyword searches on Dubawa's and AFP Fact Check Nigeria's published databases for the 2023 election cycle, supplemented by the researchers' own social media monitoring protocol. Artefacts were included if they were: (a) generated or significantly modified by AI tools; (b) political in content; and (c) targeted at Nigerian voters during the election campaign period. Phase 2 interviewees were purposively sampled across four categories (Table 2). Phase 3 survey respondents were recruited through a stratified random sampling protocol using Nigeria's electoral ward structure, with proportional allocation to each geopolitical zone.

3.8 Inclusion and Exclusion Criteria

Inclusion (Phase 1 artefacts):

- Content verified as AI-generated or AI-significantly-modified by at least one accredited fact-checking organisation
- Political content directly referencing candidates, parties, INEC, or the 2023 election
- Content circulated within Nigerian digital platforms (WhatsApp, Facebook, X/Twitter, TikTok, Instagram) between October 2022 and March 2023
- Content available in English, Hausa, Yoruba, Igbo, or Nigerian Pidgin

Exclusion (Phase 1 artefacts):

- Content where AI involvement could not be verified by a credible source
- Content originating from sources outside Nigeria or primarily targeting non-Nigerian audiences
- Satirical content clearly labelled as such at point of original publication

Inclusion (Phase 3 survey respondents):

- Nigerian citizen aged 18 years or above
- Registered voter on the INEC register as at the 2023 general election
- Mobile phone access (feature or smartphone) at the time of the 2023 election

Exclusion (Phase 3 survey respondents):

- Persons with no recall of the 2023 election campaign period
- Persons employed by political parties, INEC, or technology companies in roles directly related to the 2023 election

3.9 Procedure and Methodology

Phase 1 artefacts were coded by a team of four trained research assistants using a codebook developed from the literature

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria’s Multi-Ethnic Digital Space

and refined through iterative pilot coding on a 10% sub-sample (n=124). Inter-rater reliability was assessed using Cohen’s kappa ($\kappa = 0.83$, indicating strong agreement). Artefacts were coded for: manipulation modality (audio deepfake, video deepfake, AI-generated text, AI-amplified bot content, composite); primary language; ethnic/regional targeting (Hausa, Yoruba, Igbo, Ijaw, other, non-ethnic); political intent (negative candidate portrayal, positive candidate portrayal, institutional delegitimisation, voter suppression, ethnic incitement); and platform of primary circulation.

Phase 2 interviews were conducted by the lead researcher and one research assistant, audio-recorded with consent, transcribed verbatim, and translated where necessary by certified translators. Analysis followed Braun and Clarke’s [28] thematic analysis framework, supported by NVivo 14 software. Phase 3 survey instruments were administered via USSD (for feature phone users) and mobile web, in English, Hausa, Yoruba, Igbo, and Nigerian Pidgin, with translation verified by native speaker review panels for each language version.

3.10 Statistical Analysis

Descriptive statistics characterised the distribution of artefact types, targeting patterns, platform circulation, and voter exposure. Chi-square tests assessed associations between geopolitical zone and deepfake exposure. Binary logistic regression modelled predictors of self-reported belief change following AI-generated content exposure. Thematic analysis of interview data was triangulated with quantitative findings. All analyses were conducted using SPSS v29 and R (version 4.3.2). The significance threshold was set at $p < 0.05$.

3.11 Ethical Considerations

All interview participants provided written informed consent in their preferred language. Survey respondents were fully anonymised through code substitution. Content analysis artefacts were analysed as public-domain digital content, consistent with established social media research ethics guidelines. The research team took specific precautions to avoid republishing or amplifying harmful synthetic content in the course of analysis.

4. Results

4.1 Profile of Identified AI-Mediated Electoral Manipulation Artefacts

Content analysis of the 1,240 identified artefacts revealed a clear taxonomy of AI-mediated electoral manipulation modalities. Table 1 presents the distribution by manipulation type, with corresponding prevalence and dominant platform of circulation.

AMEM Type	Count (n)	% of Corpus	Primary Platform	Dominant Target
AI-generated audio deepfake	312	25.2	WhatsApp voice notes	Ethnic/religious identity
Video deepfake (face/voice swap)	228	18.4	TikTok, Facebook	Candidate reputation
AI-generated text disinformation	381	30.7	X (Twitter), Facebook	All ethnic groups
Bot-amplified synthetic content	196	15.8	X (Twitter), Instagram	Youth voter blocs
Composite (multi-modal AI content)	123	9.9	WhatsApp, Telegram	Regional voter blocs
TOTAL	1,240	100.0	—	—

Table 1: Distribution of AI-Mediated Electoral Manipulation (AMEM) Artefacts by Type, 2023 Nigerian Election Cycle (n=1,240).

AI-generated text disinformation was the most prevalent modality (30.7%), consistent with the comparatively low technical barrier to text-based synthetic content generation. AI-generated audio deepfakes were the second most prevalent (25.2%) and were identified as the most contextually damaging modality by Phase 2 interviewees, given the dominant role of WhatsApp voice notes in political communication among lower-literacy audiences. A Phase 2 interviewee from a leading Nigerian fact-checking organisation articulated the operational challenge:

“With text content we have keyword detection systems, we have Google reverse image search, we have a community of journalists looking at content. But a voice note in Hausa or Igbo sent person-to-person through WhatsApp? By the time it reaches us, it has already been forwarded five hundred times. There is no system. There is no early warning.” (Phase 2 interviewee, fact-checking organisation, Lagos)

4.2 Ethnic and Regional Targeting Patterns

Ethnic and religious targeting was identified as the dominant manipulation intent, present in 71.3% of all artefacts. Table 2 presents the ethnic targeting distribution of the artefact corpus, disaggregated by manipulation modality.

AMEM Modality	Hausa Target (%)	Yoruba Target (%)	Igbo Target (%)	Ijaw/Other (%)	None (%)
Audio deepfake	44.2	31.7	18.3	4.2	1.6
Video deepfake	28.5	37.4	27.6	3.1	3.4
AI-generated text	31.2	28.8	33.4	4.8	1.8
Bot-amplified	22.4	31.6	29.8	7.3	8.9
Composite	38.2	29.3	26.8	4.1	1.6
All artefacts (weighted)	33.4	31.7	26.2	4.7	4.0

Table 2: Ethnic Targeting Distribution of AMEM Artefacts by Modality (n=1,240).

Audio deepfakes showed the highest rate of Hausa-language targeting (44.2%), consistent with the dominant role of voice communication among Hausa-speaking audiences and the electoral significance of the North-West geopolitical zone. AI-generated text content showed a more balanced ethnic distribution, reflecting its deployment across multiple platforms and the relative ease of multi-language text generation. Critically, no artefact explicitly identified its own AI provenance; all were designed to be received as authentic organic content.

4.3 Voter Survey Results: Exposure, Belief, and Institutional Trust

Table 3 presents key findings from the Phase 3 voter survey (n=720), disaggregated by geopolitical zone and by digital literacy level.

Metric	NW (%)	NE (%)	NC (%)	SW (%)	SE (%)	SS (%)
Encountered AI political content	71.4	62.3	65.8	74.2	69.1	61.7
Believed content was authentic	54.3	61.8	49.2	38.7	42.6	57.4
Changed vote preference after exposure	18.6	22.4	16.3	11.8	14.2	19.7
Reported content to platform/authority	4.2	3.1	5.6	8.4	6.3	3.8
Lost trust in INEC after exposure	61.2	68.4	58.7	49.3	52.1	66.3
Lost trust in all news media	47.8	53.2	44.6	38.4	41.7	51.2

Table 3: Voter Survey Results by Geopolitical Zone (n=720).

Overall, 67.4% of surveyed registered voters reported encountering AI-generated political content in the 30 days preceding the 2023 election. Belief rates were notably higher in the North-East (61.8%) and North-West (54.3%) zones, where digital literacy scores were lowest on average. The impact on institutional trust was striking: across all zones, a majority of respondents who encountered AI-generated political content subsequently reported reduced trust in INEC (weighted mean: 59.3%), and a near-majority reported reduced trust in all news media (weighted mean: 46.2%). This finding is consistent with Vaccari and Chadwick’s [21] theory of deepfake uncertainty: the primary electoral effect of synthetic media is not necessarily to persuade voters of specific false claims, but to generate a generalised epistemic crisis that undermines confidence in democratic institutions.

4.4 Predictors of Belief in AI-Generated Electoral Content

Binary logistic regression modelling was used to identify significant predictors of self-reported belief in AI-generated political content among survey respondents. Table 4 presents the model results.

Predictor Variable	B	OR (95% CI)	Wald	p-value
Low digital literacy (ref: high)	1.24	3.46 (2.21–5.42)	27.81	<0.001
Feature phone only (ref: smartphone)	0.98	2.66 (1.74–4.07)	18.43	<0.001
Age 45+ (ref: 18–34)	0.87	2.39 (1.58–3.60)	15.27	<0.001
No tertiary education (ref: tertiary)	0.76	2.14 (1.42–3.22)	11.84	0.001
North-East zone (ref: South-West)	0.64	1.90 (1.23–2.93)	8.12	0.004
North-West zone (ref: South-West)	0.54	1.71 (1.12–2.63)	6.24	0.012
Hausa language primary (ref: English)	0.49	1.63 (1.07–2.49)	5.11	0.024
Gender (female, ref: male)	0.18	1.20 (0.84–1.71)	1.12	0.290

Table 4: Binary Logistic Regression — Predictors of Belief in AI-Generated Electoral Content (n=720). $R^2=0.41$ (Nagelkerke).

Low digital literacy was the strongest predictor of belief in AI-generated electoral content (OR=3.46, $p<0.001$), followed by feature phone-only access (OR=2.66), age 45 and above (OR=2.39), and the absence of tertiary education (OR=2.14). Geographic zone remained significant after controlling for these individual-level factors, with North-East (OR=1.90) and North-West (OR=1.71) respondents significantly more likely to believe AI-generated content than South-West respondents. Gender was not a significant predictor after controlling for other variables.

4.5 Regulatory Gap Analysis

Phase 2 interviews with regulatory and civil society informants, triangulated with documentary analysis of Nigeria's legal and regulatory architecture, identified five structural governance gaps. Table 5 presents the gap analysis matrix.

Regulatory Dimension	Current Instrument	Identified Gap	Consequence for Electoral Integrity
Synthetic media prohibition	Cybercrimes Act 2015, Electoral Act 2022	Neither instrument defines or prohibits AI-generated synthetic political media	No legal basis to remove or prosecute deepfake electoral content
Platform accountability	No specific instrument	No take-down obligation, timeline, or penalty for platforms hosting AI electoral manipulation	Deepfake content persists after identification; platforms self-regulate inconsistently
Pre-poll detection	INEC AI Division (est. May 2025)	No technical detection capacity for multilingual audio deepfakes; INEC AI Division underfunded	Audio deepfakes on WhatsApp undetected until post-election debunking
Digital literacy	Nigeria National Digital Literacy Framework	Framework not operationalised at ward level; no electoral-specific AI literacy component	Citizens unable to identify synthetic content; belief rates remain high
Cross-border coordination	No instrument	No mechanism to coordinate takedown of AI content originating from outside Nigeria	Foreign-hosted disinformation campaigns immune to domestic enforcement

Table 5: Regulatory Gap Analysis for AI-Mediated Electoral Manipulation in Nigeria.

5. The Democratic Consent Protection Framework (DCPF)

5.1 Framework Rationale and Design Principles

The Democratic Consent Protection Framework (DCPF) is proposed as a four-pillar governance architecture for the protection of democratic consent from AI-mediated electoral manipulation in Nigeria's multi-ethnic digital space. The framework is grounded in four design principles derived from the empirical findings. First, speed-centricity: given that the

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria’s Multi-Ethnic Digital Space

primary harm of AI electoral manipulation occurs within hours of content release, governance mechanisms must operate on electoral time — measured in hours, not weeks. Second, linguistic specificity: any detection, removal, and counter-messaging system must operate in Hausa, Yoruba, Igbo, and Nigerian Pidgin, not only in English. Third, institutional plurality: no single institution like INEC, NITDA, or any platform possesses the capacity to address AI electoral manipulation alone; effective governance requires coordinated multi-actor response. Fourth, proportionality: the framework must protect democratic speech while targeting only AI-generated synthetic media designed to deceive voters. Table 6 presents the DCPF architecture.

Pillar	Name	Core Mechanism	Primary Actor	Timeline
Pillar 1	Synthetic Media Detection	Multilingual AI deepfake detection unit within INEC AI Division; real-time monitoring of WhatsApp, Facebook, X, TikTok in Hausa, Yoruba, Igbo, Pidgin	INEC AI Division + NITDA	Operational before 2027
Pillar 2	Platform Accountability Regime	Mandatory Electoral Content Code: 48-hour take-down obligation for verified AI electoral manipulation; N50m per violation penalty; pre-election transparency reporting	FCCPC + NCC + Platforms	Legislative by 2026
Pillar 3	Electoral AI Literacy Campaign	Ward-level AI content recognition training embedded in voter education; USSD-based fact-check service in four languages; religious and community leader briefings	INEC + Civil Society + NUC	Ongoing from 2025
Pillar 4	Cross-Border Coordination Mechanism	Nigeria-led West African Electoral AI Governance Compact through ECOWAS; intelligence sharing on foreign-origin disinformation operations; joint takedown protocols	MOFA + NITDA + ECOWAS	Compact by 2026

Table 6: Democratic Consent Protection Framework (DCPF) — Four-Pillar Architecture.

5.2 Pillar 1: Synthetic Media Detection

The first and most technically demanding pillar of the DCPF addresses the fundamental detection gap: the absence of any systematic capacity to identify AI-generated electoral manipulation in Nigeria’s dominant political communication languages in near-real time. The proposed Multilingual Electoral AI Detection Unit (MEADU) would be housed within INEC’s AI Division and would operate a continuous monitoring protocol covering WhatsApp (via voluntary share forwarding partnerships), Facebook, X, TikTok, and Telegram. Detection would employ a combination of: content-based methods (facial and audio artifact analysis for deepfake identification); metadata analysis (provenance and creation date anomaly detection); and linguistic analysis (detection of AI-generated text through perplexity and burstiness metrics in Nigerian languages). A rapid alert protocol would trigger within two hours of identification, initiating the platform take-down process under Pillar 2.

5.3 Pillar 2: Platform Accountability Regime

The second pillar addresses the absence of enforceable obligations on digital platforms operating in the Nigerian electoral space. The proposed Electoral Content Code, to be enacted as a Schedule to the Electoral Act through amendment ahead of 2027, would establish: a 48-hour mandatory take-down obligation for AI-generated synthetic political content verified by MEADU or an accredited fact-checking organisation; a pre-election content transparency reporting requirement for major platforms (monthly disclosure of electoral content flagged, reviewed, and removed); a huge financial fine per verified-violation penalty, escalating to licence suspension for repeat infractions; and a safe harbour provision protecting platforms that respond within the 48-hour window and cooperate with INEC investigation. This pillar draws on the operational experience of the EU’s Digital Services Act and Kenya’s Draft Digital Media Regulation Framework.

5.4 Pillar 3: Electoral AI Literacy Campaign

The third pillar recognises that detection and removal, however effective, cannot alone protect democratic consent in an environment where 67.4% of voters were exposed to AI electoral manipulation in a single election cycle. Voter-level resilience requires electoral AI literacy: the capacity to recognise synthetic media, to seek verification, and to withhold belief pending confirmation. The proposed ward-level AI literacy programme would be integrated into INEC’s existing voter education infrastructure, delivered through community town halls, religious organisation briefings (given the trusted status of Islamic and

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space

Christian community leaders in Nigerian political communication), and a USSD-accessible 'FactCheck' service enabling users to submit voice notes or text for automated credibility assessment in Hausa, Yoruba, Igbo, and Pidgin. The National Universities Commission (NUC) would be engaged to integrate AI media literacy into tertiary curricula.

5.5 Pillar 4: Cross-Border Coordination

The fourth pillar addresses the jurisdictional limitation that makes purely domestic governance insufficient: AI electoral manipulation operations are frequently hosted on foreign servers, coordinated across international boundaries, and technically immune to domestic enforcement action. The proposed Nigeria-led West African Electoral AI Governance Compact, to be negotiated through ECOWAS' Democracy and Good Governance Protocol structures, would establish: a shared early warning database of known AI electoral manipulation tactics, actors, and content signatures; reciprocal take-down agreements covering content targeting any ECOWAS member state's elections; joint technical capacity building for multilingual detection in West African languages; and a common definition of AI-mediated electoral manipulation for purposes of mutual legal assistance.

5.6 Framework Validation: Delphi Results

The DCPF was validated through a two-round Delphi process with a panel of 22 experts drawn from electoral governance (n=6), AI and cybersecurity (n=5), Nigerian civil society and journalism (n=5), legal and regulatory expertise (n=4), and comparative African democracy studies (n=2). Consensus was defined at $\geq 70\%$ agreement. Table 7 presents Delphi results.

DCPF Component	Round 1 Agreement (%)	Round 2 Agreement (%)	Consensus	Key Expert Concern (Round 1)
Multilingual detection unit (MEADU)	74	89	Yes	Technical capacity and funding sustainability
48-hour platform take-down obligation	68	82	Yes	Risk to free speech; need for independent verification
Platform penalty regime (N50m)	61	74	Yes	Penalty level may be insufficient for global platforms
Electoral Content Code (legislation)	71	86	Yes	Legislative timeline ahead of 2027
Ward-level AI literacy programme	82	93	Yes	Resource allocation and trainer capacity
USSD FactCheck service	79	91	Yes	Accuracy and latency of automated assessment
Religious leader briefings	74	88	Yes	Risk of partisan instrumentalisation
ECOWAS cross-border compact	66	77	Yes	Political will among ECOWAS member states

Table 7: Delphi Validation Results for DCPF (n=22 experts).

All eight DCPF components achieved consensus ($\geq 70\%$) in Round 2. The ward-level AI literacy programme received the highest consensus scores across both rounds (82% and 93% respectively), reflecting broad expert agreement that civic education is the most sustainable and politically uncontroversial governance intervention. The ECOWAS cross-border compact achieved the lowest Round 2 consensus (77%), reflecting genuine uncertainty about political will among member states; panellists recommended that Nigeria negotiate the compact bilaterally with Ghana and Senegal as priority partners before seeking regional multilateral agreement.

6. Discussion

The findings of this study establish, empirically and analytically, that AI-mediated electoral manipulation in Nigeria is not an emerging threat to be prepared for — it is an active harm that has already shaped a national election and is escalating in both capability and scale ahead of 2027. The 67.4% exposure rate, the 59.3% institutional trust erosion rate, and the 71.3% ethnic targeting rate document not merely a technology policy problem but a democratic crisis with immediate and measurable

AI-Mediated Electoral Manipulation and the Fragility of Democratic Consent in Nigeria's Multi-Ethnic Digital Space

consequences for the quality of consent underlying Nigeria's electoral outcomes.

The study's most theoretically significant finding is the confirmation and quantification of Vaccari and Chadwick's [21] deepfake uncertainty thesis in a Sub-Saharan African multi-ethnic context. The primary electoral harm documented in this study is not that a significant proportion of voters were persuaded to vote for the 'wrong' candidate by synthetic content — though the 14.2–22.4% zone-level vote preference change rates are significant. The primary harm is the erosion of epistemic trust: the capacity for citizens to distinguish authentic from synthetic political information, and the confidence in democratic institutions — most critically INEC — required to regard electoral outcomes as legitimate. This finding has constitutional implications: if AI-mediated manipulation systematically corrupts the informational conditions of democratic consent, the legitimacy of electoral outcomes becomes structurally questionable, regardless of the procedural regularity of the vote count itself.

The ethnic targeting dimension of this corruption is particularly acute in the Nigerian context. Nigeria's political history is characterised by the strategic mobilisation of ethnic identity as a voting cue [15, 16], and the country's democratic stability rests on a complex system of informal ethnic power-sharing arrangements that have never been fully constitutionalised [29]. AI systems capable of generating ethnically targeted synthetic content in Hausa, Yoruba, and Igbo at negligible cost introduce a qualitatively new instrument of ethnic political manipulation — one that can simultaneously reinforce in-group identity and manufacture out-group threat narratives across all three major ethnic blocs from a single operational command. The Dubawa [4] documentation of a coordinated ethno-disinformation network operating before and during the 2023 election provides real-world evidence of this threat.

The DCPF framework responds to this complexity with a multi-actor, speed-centric, linguistically specific governance architecture. Its novelty lies not in any single mechanism — deepfake detection technology, platform accountability, voter education, and cross-border coordination are each discussed in the existing literature — but in the integration of these elements into a framework specifically calibrated to Nigeria's multi-ethnic political economy and its specific digital communication environment. The USSD FactCheck service, for instance, is designed precisely for the 41.5% of North-West zone respondents who access the internet only through feature phones and who consume political content primarily through WhatsApp voice notes in Hausa. No existing global AI electoral governance framework accommodates this user profile.

Several limitations of the study must be acknowledged. The content analysis corpus, though comprehensive by available standards, is necessarily incomplete: WhatsApp-distributed content that was never fact-checked by an accredited organisation is, by definition, absent from the sample. This introduces a systematic under-count of the most difficult-to-detect and most widely distributed manipulation modality — audio deepfakes on closed messaging platforms. Additionally, the DCPF's Pillar 2 (platform accountability) faces implementation risk given the limited operational presence of Meta, TikTok, and X in Nigeria, and their historical reluctance to subject themselves to national content regulation in African markets.

7. Conclusion

Nigeria stands at a democratic inflection point. The 2023 general election demonstrated, for the first time in any African democracy, the deployment of generative AI as a systematic instrument of ethnic political manipulation at national scale. The 2027 election will confront these threats at significantly greater technical sophistication, lower production cost, and higher penetration — and Nigeria's current governance architecture is structurally unequipped to address them. The question is not whether AI-mediated electoral manipulation will shape Nigeria's 2027 elections; on current trajectories, it will. The question is whether Nigeria will build, before 2027, the governance capacity to contain the damage to democratic consent.

This paper makes three contributions toward that capacity. Empirically, it provides the first systematic typology of AI-mediated electoral manipulation in Nigeria, documenting the distribution, ethnic targeting patterns, and voter impact of 1,240 AI-generated artefacts from the 2023 election cycle. Analytically, it demonstrates that the primary harm of AI electoral manipulation in Nigeria is not vote-switching but institutional trust erosion — a finding with profound implications for how governance responses are designed and evaluated. Normatively, it proposes and validates the Democratic Consent Protection Framework: a four-pillar, speed-centric, linguistically specific governance architecture designed for Nigeria's multi-ethnic digital space and scalable to comparable democratic contexts across West Africa.

Democratic consent cannot be secured by technology alone, nor by regulation alone, nor by education alone. It requires the simultaneous operation of all three, coordinated across institutions and calibrated to the specific vulnerabilities of the communities most at risk. In Nigeria, those communities are predominantly Hausa, Yoruba, and Igbo speakers with limited digital literacy and high exposure to WhatsApp-mediated political communication — the exact population that AI-generated audio deepfakes are designed to exploit. The DCPF is built for them.

Acknowledgements

The authors gratefully acknowledge the technical assistance of Dubawa Nigeria, AFP Fact Check Nigeria, and the Peoples' Gazette, whose publicly available fact-checking databases formed the basis of the Phase 1 content analysis corpus. The authors also thank the 22 Delphi panel experts for their rigorous engagement with the DCPF and the anonymous peer reviewers whose comments strengthened the manuscript. All errors of fact and interpretation remain the authors' own.

Conflict of Interest: The authors declare no conflict of interest.

Data Availability: The content analysis codebook, survey instrument, and Delphi framework documentation are available from the corresponding author on reasonable request for replication and pedagogical purposes.

References

1. Ajakaiye H. Q&A: Manipulated media in the 2023 Nigerian presidential elections, generative AI, and possible interventions. Institute for Security and Technology; 2024. Available: securityandtechnology.org
2. Journalistsresource.org. How AI-generated deepfakes threaten the 2024 election. Harvard Shorenstein Center; 2024. Available: journalistsresource.org
3. Africa in Fact. The challenges of AI-driven political disinformation in Nigeria. Good Governance Africa; 2024. Available: africainfact.com
4. Dubawa. Analysis: Social media users deploy ethnic hate speech, disinformation ahead of Nigeria's 2027 elections. Dubawa; 2025. Available: dubawa.org
5. Oyediji O. AI in Nigerian Elections: Tool for Safeguarding Democracy or Weapon for Deepfakes and Disinformation. *Nepal J Multidiscip Res*. 2025;8(3):89–108.
6. Techpoint Africa. AI could disrupt Nigeria's 2027 election and the country is not ready. Techpoint Africa; 2026. Available: techpoint.africa
7. TheCable. The invisible AI threat to Nigeria's 2027 elections. TheCable; 2026. Available: thecable.ng
8. Legit.ng. How malicious AI networks will threaten Nigeria's 2027 elections. Legit.ng; 2026. Available: legit.ng
9. Federal Republic of Nigeria. Electoral Act, 2022. Lagos: Government Printer; 2022.
10. Federal Republic of Nigeria. Cybercrimes (Prohibition, Prevention, Etc.) Act, 2015. Abuja: Government Printer; 2015.
11. BusinessDay. Artificial intelligence, electoral integrity, and the risk of manipulation of Nigeria's 2027 general elections in perspective. BusinessDay; 2026. Available: businessday.ng
12. Rawls J. A Theory of Justice. Cambridge: Harvard University Press; 1971.
13. Habermas J. The Theory of Communicative Action. Vol. 1. Boston: Beacon Press; 1984.
14. Zuboff S. The Age of Surveillance Capitalism. New York: PublicAffairs; 2019.
15. Horowitz DL. Ethnic Groups in Conflict. Berkeley: University of California Press; 1985.
16. Brass PR. Ethnicity and Nationalism: Theory and Comparison. New Delhi: Sage; 1991.
17. Dubawa. AI-generated disinformation looms ahead of Nigeria's 2027 elections. Dubawa; 2025. Available: dubawa.org
18. TechCabal. Preparing for Nigerian elections in a world of AI deepfakes. TechCabal; 2025. Available: techcabal.com
19. Journalistsresource.org. Global deepfake elections roundup. Harvard Shorenstein Center; 2024.
20. The Nation. Malicious AI swarms and 2027 elections. The Nation Newspaper; 2026. Available: thenationonline.net
21. Vaccari C, Chadwick A. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Soc Media Soc*. 2020;6(1):2056305120903408.
22. Wardle C, Derakhshan H. Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Strasbourg: Council of Europe; 2017.
23. Bradshaw S, Howard PN. The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. Oxford Internet Institute; 2019.
24. Okoro U, Mackenzie T, Radeljic B. Governance and Regulation of Artificial Intelligence in Developing Countries: A Case Study of Nigeria. *arXiv*. 2026; 2604.06018.
25. Pasipamire N, Muroyiwa A. Navigating algorithm bias in AI: ensuring fairness and trust in Africa. *Front Res Metr Anal*. 2024. doi:10.3389/frma.2024.1486600.
26. NITDA. Nigeria National AI Policy 2024. Abuja: National Information Technology Development Agency; 2024.
27. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. 3rd ed. Thousand Oaks: Sage; 2018.
28. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77–101.
29. International IDEA. Addressing Nigeria's Unresolved Electoral Reform Issues. Stockholm: IDEA; 2024. Available: idea.int
30. DataReportal. Digital 2024: Nigeria. DataReportal; 2024. Available: datareportal.com