

---

# **Artificial Intelligence-Enabled Disinformation And Electoral Security In Nigeria: A Critical Review Of Emerging Threats, Regulatory Gaps And Policy Responses**

Review Article | Volume 1 | Issue 4 | 2026 | Article Number: 030

Accepted: 29 August 2026 | Published: 10 September 2026 | ISSN: 2979-8582 (Online)



Crossref : <https://doi.org/10.67693/BJCR-R75MWTk5>



---

**Sakeena Audu**

Federal University of Lafia, Nigeria

---

Correspondence: Sakeena Audu, [Sakeena.audu2020@nda.edu.ng](mailto:Sakeena.audu2020@nda.edu.ng)

---

## **Abstract**

Generative artificial intelligence has transformed the production of political disinformation, lowering the cost of fabricating convincing synthetic audio, video, and text to nearly zero and enabling forms of automated, personalised manipulation that older propaganda techniques could not sustain. This article offers a critical review of what this shift means for electoral security in Nigeria, a democracy already marked by ethno-religious fragmentation, a winner-takes-all political culture, and uneven trust in electoral institutions. Drawing on conceptual and theoretical literature, empirical evidence from Nigeria's 2023 general election, and an assessment of the regulatory and policy landscape, the review argues that AI functions less as a novel political actor than as a force multiplier for disinformation practices that predate it, gaining its electoral significance from how it interacts with, rather than replaces, existing Nigerian vulnerabilities. Documented incidents from the 2023 election, including deepfake endorsements, fabricated audio alleging electoral rigging, and the liar's dividend invoked around a genuine campaign gaffe, illustrate how synthetic content exploited institutional weaknesses such as the technical failures of the Bimodal Voter Accreditation System and the IReV results portal. The review finds Nigeria's regulatory response reactive and fragmented, constrained by definitional gaps, overlapping institutional mandates, weak technical verification capacity, and an unresolved tension between restricting harmful content and protecting legitimate political speech. It concludes that closing this gap requires Nigeria-specific empirical research on actual voter response, longitudinal and comparative African evidence, and interdisciplinary work connecting technological, institutional, and legal capacity, since no single one of these alone accounts for how AI-enabled disinformation threatens electoral security.

**Keywords:** Artificial Intelligence, Disinformation, Electoral Security, Deepfakes, Nigeria, Ai Governance, Democratic Resilience

## 1.0 INTRODUCTION

The wake of the internet era brought out heightened optimism across the globe, up until the point of decades of ‘field-testing,’ which now poses a significant and systematic democratic threat. Particularly in 2016, a consensus of academic and public perception regarding the impact of digital media on democracy was born, following the political drama that ensued from the 2016 US presidential election and the UK’s Brexit referendum. More so, contrary to the expected “utopianism” tagged on the advent of social media and “Information revolution,” the outcomes have been marred by its capacity for mass manipulation, along with other negative shifts recorded (Alberta, 2025). These shifts from global optimism to global concern were well captured by Lupin (2020), who described the post-2016 reckoning of the internet as an ushering into illiberal ends.

The core issues with the current state of global information can be summed up as a “systemic disinformation order”. Disinformation in this context implies a systemic disruption of authoritative information flow (Livingston, 2018), hence jeopardizing the outlook of global democracy and institutional trust. Lupin (2020) also bemoans the uprising of “indeterminacy culture,” where what is held as collective truth is already compromised from its fundamentals. The decline in reasoned debate practices in public over social issues has further widened the gap in information determinacy, thereby subjecting all to shared truths from untrusted institutions.

Furthermore, the rise of computational propaganda and weaponized media algorithms has exacerbated the deception by industrializing its scope while leveraging automation to simulate the process. Hence, as a means of a political tool, sentiments are fed to the public at scale through surveillance technologies and rapidly coordinated bot networks (Olawunmi, 2015). Beyond the level of automated algorithms, manipulation of public opinions is also taken further by the professionalization of “disinformation as a service” through the hiring of private consultants to push a public agenda (Mansur, 2025). These all set the precedent for the adoption of AI in specific content generation with little or no scrutiny on the nobility of intent or authenticity of the message passed.

In modern times, the proliferation of hyper-realistic synthetic media has made the entire scenery of indeterminacy even worse. According to Vaccari & Chadwick (2020), it induced an acute state of “climate of ‘indeterminacy’ about truth and falsity”, thereby blurring the lines between what is truth from fiction. When these deepfakes are used in political spheres, electorates are constantly thrown into confusion, as they cannot differentiate truth from fabrication (Mansur, 2025). This environment fosters a broader condition of transparency subversion where malicious political actors directly or indirectly take advantage of the possibilities of deepfakes. By direct impact, they claim and align with every AI generation that favours their campaign, and indirectly, they distance themselves from every media publication that criticizes them, claiming it is AI-generated, irrespective of whether those publications are real or not. This behooves this synthetic era to defend the truth by shifting the burden of proof (Pawalec, 2022). Furthermore, the algorithm of social media platforms deepens this decay by prioritizing emotionally charged and polarised content to maximize engagement (Olawunmi, 2025). Hence, users are constantly trapped within digital media environments that constantly reinforce existing bias. The final result is a state of pernicious polarization where opposing political groups view each other as illegitimate, thereby dampening every hope of rational democratic deliberation (Vasist et al., 2023). While these epistemic and structural threats are global, they manifest with a unique and deeper level of intensity in Nigeria, which is already an ethnically fragmented nation with several elements of institutionalised divisions.

### 1.1 Socio-Political Fragmentation in Nigeria: The 2023 Watershed

The root of Nigeria’s susceptibility to AI-driven manipulation is linked to existential ethno-religious fault

lines of division. In addition, the political paradigm in the country, which assumes a zero-sum nature, described by Okezie & Inobemhe (2025) as a “winner-takes-all or do-or-die political mindset,” further worsens the vulnerabilities. As a result, the stakes of most political contests are hardly driven by ideological, but rather, a quest for ethnic and regional dominance. In the 2023 Nigerian general elections, there was a scaled use of deepfake endorsements and fabricated audios within the political scene (Ekpang et al., 2023). Notable examples include synthetic videos of political actors particularly showing Hollywood stars endorsing presidential candidate Peter Obi (Ekpang et al., 2023). Also, synthetic audios were used to incite tension as they purported “Atiku/Tambuwal/Okowa rigging plans” (Okezie & Inobemhe, 2025). These deceptions were strategically deployed as tools for destabilizing the polity and portraying the actors involved as “misfits” for political roles (Ekpang et al., 2023). The overarching impacts of these negative media strategies are reported by the Center for Democracy and Development to have influenced the voting choices of 15% of the responding electorates (Albert, 2025). While a conclusive claim on the impacts cannot be reached at this point, it remains undeniable that the speed and spread of synthetic narratives of generative AI pose serious concerns and are aided by digital platforms.

In the Nigerian context, many observers have identified that WhatsApp acts as the primary “battlefield” for disinformation, firstly due to its reach. Also, Olawunmi (2025) noted that the platform’s end-to-end encryption protects the actors of disinformation from external monitoring and renders debunking exceptionally difficult. Unsubstantiated claims were quickly disseminated through WhatsApp and other encrypted messaging platforms exploiting the high degree of trust on the part of users in their personal networks (Albert, 2025). However, the wave of disinformation and digital manipulation reached a tipping point in the 2023 elections where technical malfunctions in the “Bimodal Voter Accreditation System (BVAS)” and the “INEC Results Viewing (IREV) portal” provided “digital evidence” to rigging narratives (Asangausung et al., 2025). The failure to upload results in real time fueled suspicions and electorates started to question the validity of the process. Thus, the ensuing noise was rapidly amplified by AI-driven systems to delegitimize the outcome of the electoral process (Asangausung et al., 2025). This intersection between coordinated lies and institutional failure created a climate of doubts that discouraged voter participation and undermined public confidence in the Independent National Electoral Commission (INEC). Had Nigeria’s governance systems provided regulatory frameworks for the use of generative AI during elections, some of the resulting ills from that amplification, like physical insecurity, fear, doubt, etc., could have been brought under control.

The 2023 general elections in Nigeria recorded a low voter turnout of only 26.7% (Namadi & Yerima, 2025). However, the escalation of AI-generated lies and disinformation from the buildup led to a spike of violence which had over at least 20 INEC in flames (Oni et al., 2025). This insecurity highlights a profound gap in regulation which aided the rapidly evolving AI technology to outpace the laws that currently govern electioneering in the country. Even globally, the governance systems are still fragmenting, allowing malicious actors to exploit loopholes. While legislative efforts such as the Electoral Act, 2022 and the 2023 Social Media Bill exist to modernize Nigeria, they are weakly enforced and often criticized for their potential state censorship or digital authoritarianism (olawunmi, 2025). This policy vacuum necessitates a transition toward an Integrated resilience framework. However, it is noteworthy that the current scholarship on AI disinformation remains largely far from being holistic. Leaving a critical gap regarding the long-term impact of AI on Nigerian electoral integrity and the effectiveness of local policy responses. Against this backdrop, this review addresses this need by pursuing three primary objectives: first, evaluating the emerging threats of AI-enabled content; second, identifying the regulatory and institutional gaps that allow these threats to proliferate; and third, proposing an integrated resilience framework for Nigeria’s future electoral security.

## 2. CONCEPTUAL AND THEORETICAL FOUNDATIONS

### 2.1 Conceptualising Disinformation

Research on digital manipulation has moved away from treating “fake news” as an isolated pathology and toward a more structural account of what Bennett and Livingston (2018) term the “disinformation order,” a condition in which the routine disruption of authoritative information flows becomes a durable feature of political communication rather than an occasional aberration. Within this literature, three categories are conventionally distinguished. Misinformation refers to false or misleading content shared without an intent to deceive; disinformation describes falsehoods deliberately fabricated or distorted for a strategic purpose; and malinformation covers accurate information deployed maliciously, as in the selective leaking of private material to inflict reputational harm. The boundary that matters most for this review is intentionality. Disinformation is defined here as information that is knowingly false or manipulated and strategically deployed to alter perception, discredit opponents, or manufacture an appearance of consensus that would not otherwise exist. This intentional character is what makes disinformation a political instrument rather than a byproduct of a noisy information environment.

Locating disinformation within political communication requires treating it as part of a broader influence process rather than a discrete artefact. A single fabricated audio clip or manipulated image rarely does the work of disinformation on its own; its effect depends on the networks, platforms, and audiences that carry it, and on the pre-existing grievances it activates. This is a sort of algorithmic manipulation, as Mansur (2025) writes, where technology is used less to produce new political motives, than to industrialize the pursuit of existing ones. Olawunmi (2025) situates disinformation within the phenomenon of computational propaganda where automation and coordination enable small actors to manufacture the appearance of mass sentiment.

Scholars disagree about the conceptual weight to be assigned to any one technology. Others caution against overstating deepfakes, which they contend detracts attention from the broader domain of generative propaganda where AI is employed for persuasive and promotional ends that stop short of outright fabrication but still distort public reasoning (Beacken et al., 2026). Others warn against what might be called technochauvinism, the tendency to treat AI as an unprecedented epistemic rupture while downplaying the long history of propaganda and the social conditions that make audiences receptive to it. Much of the conceptual framing of disinformation research is rooted in the Global North and assumes open platform-based information ecosystems, as noted by Alberts (2025) and Okolo (2024). This assumption fits poorly with contexts such as Nigeria, where closed, encrypted channels like WhatsApp function as the primary vector for political rumour and where trust in personal networks often exceeds trust in formal media or institutions.

For the purposes of this review, disinformation is conceptualised as a form of computational influence operation enabled by generative AI: intentional, strategically directed information manipulation whose scale and sophistication are qualitatively altered by automated production and distribution. This framing retains the intentionality that separates disinformation from misinformation while making room for the industrial capacity that generative AI adds to older propaganda techniques.

### 2.2 AI-Enabled Disinformation and Electoral Security

What generative AI changes is less the motive behind disinformation than its production function. Large language models and generative adversarial networks now allow text, audio, image, and video content to be produced at a volume and fidelity that manual fabrication could never sustain, and AI-generated text is frequently indistinguishable from professionally written journalism (Ferrara, 2025; Shoaib et al., 2023).

Deepfakes occupy a particular place in this landscape because they operate at what Pawelec (2022) calls a visceral level, exploiting the intuitive trust audiences place in seeing and hearing a person rather than merely reading a claim about them. Automation not only shortens the time between production and dissemination, allowing synthetic content to circulate faster than fact-checking infrastructures can respond to, but also offers personalization techniques that use behavioral and psychographic data to adapt disinformation to the psychological profile of specific audiences rather than broadcasting a single message to all (Ferrara, 2025; Mansur, 2025). None of this would go far without the platforms that distribute it. Engagement-optimised recommendation systems reward emotionally charged and polarizing content, regardless of its veracity, leading to echo chambers and filter bubbles that concentrate exposure to synthetic narratives among audiences already primed to accept them (Vasist et al., 2023).

Electoral security here means not only physical safety of the polling infrastructure, but also the integrity of the information environment in which electoral choices are made. This includes the reliability of the information on which voters base their understanding of candidates and institutions, the perceived legitimacy of electoral management bodies and overall resilience of democratic decision-making to manipulation. Bennett and Livingston (2018) treat the erosion of authoritative information flows as itself a threat to democratic institutions, and this logic applies directly to elections, where disinformation targeting an electoral commission does not merely spread a false claim but corrodes the perceived legitimacy of the process the commission administers. A closely related mechanism is what Pawelec (2022) and Vaccari and Chadwick (2020) describe as a trust crisis, in which the mere existence of convincing synthetic media allows political actors to dismiss authentic evidence as fabricated, a dynamic often labelled the liar's dividend.

The relationship between AI capability and electoral harm is best understood as a pathway rather than a direct causal jump: generative capacity enables the production of synthetic disinformation, which is scaled through coordinated distribution and micro-targeting, amplified by engagement-driven algorithms, and only then converted into electoral consequences such as trust decay, polarisation, and depressed participation. This pathway is not automatic, and the evidence supporting each link varies in strength. The decline of confidence in electoral institutions and the proliferation of deepfake endorsements in Nigeria's 2023 elections are fairly well documented (Ekpang et al., 2023; Okezie et al., 2025). The use of AI-generated audio to create rigging narratives and the link between suspicion driven by misinformation and record low turnout are similarly based on observed events, even if determining direct causal weight is difficult (Asangausung et al., 2025; Namadi & Yerima, 2025). Other claims stand on less solid ground. We have documented the rise of AI-generated news outlets in sub-Saharan Africa, and the use of conversational AI to achieve individualized persuasion, but these are still early-stage phenomena, and their significance for elections is not yet settled (Okolo, 2024). At the far end sit speculative claims of an impending epistemic collapse in which the public lose the capacity to distinguish truth from fabrication altogether. This scenario recurs in the literature as a theoretical concern rather than an empirically demonstrated outcome, and treating it as settled fact would overstate what the evidence currently supports.

This distinction matters for a second reason. It bears directly on whether AI represents a genuinely new category of electoral threat or an acceleration of disinformation practices that predate it. The propaganda techniques AI now automates, fabrication, targeting, and coordinated amplification, are not new inventions. What generative AI supplies is scale, speed, and a realism that lowers the cost of producing convincing fabrications to nearly zero. Read this way, AI is less a new political actor than a force multiplier for existing political motives, a distinction that matters for how the threat is analysed in Section 3 and how it should be regulated in Section 4.



## 2.3 Theoretical Perspectives

Two complementary theoretical perspectives do most of the analytical work in this review. Institutionalism directs attention to the rules, norms, and organisational capacity of the bodies responsible for managing elections, and treats electoral integrity as substantially a function of how well those institutions perform under stress rather than as an automatic outcome of holding elections at all. Applied to Nigeria, this perspective is useful because it locates part of the 2023 electoral security failure in the technical and organisational performance of the Independent National Electoral Commission, including the reliability of the Bimodal Voter Accreditation System and the Results Viewing portal, rather than treating disinformation as an external force acting on a neutral institutional backdrop (Madueke & Enyiazu, 2025). Its limitation is that it can understate the role of factors that lie outside any single institution's control, such as platform architecture or transnational coordination.

Complexity theory addresses that gap. It treats polarisation and disinformation as emergent properties of a system in which state actors, political parties, platforms, and ordinary users interact, and it resists attributing electoral security outcomes to any single cause in isolation. Vasist et al. (2023) apply a configurational approach of exactly this kind to show that polarisation results from particular combinations of conditions rather than from disinformation exposure alone, a framing well suited to Nigeria's overlapping ethno-religious cleavages, closed messaging ecosystems, and uneven institutional capacity. Where institutionalism explains why INEC's technological and organisational performance mattered so much in 2023, complexity theory explains why the same disinformation content produced different effects across Nigeria's regional and religious fault lines.

A third perspective, democratic innovation theory, is worth acknowledging briefly for the tension it introduces rather than as a primary framework. It holds that AI can strengthen democratic processes, through automated fact-checking or transparency tools, even as the same underlying technology is weaponised against those processes (Romanishyn et al., 2025). This ambivalence is a useful corrective against treating AI as a uniformly destructive force, though it is not developed here as a central lens.

Combining these, institutionalism and complexity theory provide the analytical foundation for the argument that follows. They direct attention to institutional performance and systemic interaction rather than to technology in isolation, which is precisely the orientation needed to examine how AI-enabled disinformation intersects with Nigeria's specific electoral vulnerabilities in Section 3.

## 3. AI-ENABLED DISINFORMATION AND EMERGING ELECTORAL THREATS IN NIGERIA

### 3.1 Evolution of Electoral Disinformation

Nigerian electoral disinformation did not begin with artificial intelligence. Rumour campaigns questioning candidates' health, ethnicity, or religious affiliation have shadowed Nigerian elections since well before the advent of social media, and the 2019 cycle already saw viral falsehoods, most notably claims that President Muhammadu Buhari had died abroad and been replaced by a body double, spread through platforms with little institutional capacity to counter them. What has changed is the labour required to produce convincing falsehoods and the speed at which they travel. Loth et al. (2026) describe this shift as a movement from disinformation as a craft, dependent on manual editing and coordinated human effort, to disinformation as an industry, in which generative tools allow a single actor to produce synthetic content at a volume once associated with state-backed operations. Nigeria's 2023 general election is widely treated in the literature as the moment this industrial mode became visible in a Nigerian electoral context, with AI-generated audio and video appearing alongside the older repertoire of doctored images and recycled footage from previous elections (Okezie et al., 2025; Hassan, 2023).

This evolution is better read as an acceleration than a rupture. The underlying political motives, discrediting rivals, inflaming ethnic and religious grievance, casting doubt on electoral administration, are the same motives that have driven Nigerian political communication for decades. What generative AI supplies is a lower cost of entry and a higher ceiling of realism, allowing actors who previously lacked the technical skill or financial resources for sophisticated manipulation to produce content that rivals what once required a professional production team.

### 3.2 Emerging AI Threat Landscape

The clearest capability shift lies in synthetic media production. Generative adversarial networks and diffusion models now produce video and audio manipulations that are difficult to distinguish from authentic footage without specialised forensic tools, and large language models such as GPT-4 can generate politically persuasive text, complete with the tone and idiom of local news reporting, far faster than any human writing team (Loth et al., 2026; Islas et al., 2024). During the 2023 elections this capability produced specific, documented artefacts: a synthetic video in which international celebrities appeared to endorse Peter Obi's candidacy, and fabricated audio purporting to capture opposition figures plotting to rig the outcome (Ekpong et al., 2023; Okezie et al., 2025). The electoral security implication is not simply that false content circulated, disinformation has always circulated, but that its evidentiary form, video and voice, carries a persuasive weight that text-based rumour does not. Experimental work outside Nigeria confirms this asymmetry: visual and audio manipulations tend to produce measurably greater shifts in belief and reported voting intention than equivalent false claims presented as text, and journalistic fact-checks only partially reverse the damage once exposure has occurred (Dan, 2025; Vaccari & Chadwick, 2020).

A second and less visible threat concerns automated content production, including a new generation of social bots capable of sustaining humanlike conversation. Doshi et al. (2024) demonstrate that language models can now drive bot accounts that adapt their arguments in real time, build rapport with individual users over days or weeks, and gradually shift political sentiment through what amounts to simulated peer persuasion rather than blunt broadcast messaging. Nigeria has not yet produced documented cases of this precise technique, but the underlying tools are freely available, and the country's political communication already relies heavily on peer-to-peer trust networks, a structure this kind of bot is designed to exploit.

Microtargeting compounds both problems. Where earlier disinformation was broadcast indiscriminately, AI-assisted targeting allows a message to be tuned to the specific grievances of a religious community, an ethnic bloc, or a regional audience, exploiting exactly the cleavages that Nigerian political competition already tends to activate (Khan et al., 2026). Algorithmic amplification then does further work independent of any single actor's intent. Platform recommendation systems reward engagement, and synthetic content engineered to provoke outrage or fear tends to travel further and faster through a network than neutral factual reporting, a pattern that holds across the platforms most used for Nigerian political discussion, including WhatsApp, X, and TikTok (Loth et al., 2026).

These dynamics converge on two related consequences for electoral security. The first is what several sources term the liar's dividend: once audiences know convincing fabrications are possible, genuine and damaging evidence can be dismissed as synthetic. Nigeria has already produced a domestic example of this logic, when a campaign spokesperson for Bola Tinubu claimed a widely circulated video of an authentic verbal gaffe had been digitally doctored, a claim independent fact-checkers found to be false (Okezie et al., 2025). Deepfakes need not even be used to erode trust; the mere plausibility of their use is sufficient. The second consequence is a broader epistemic fatigue, sometimes labelled epistemic cynicism, in which repeated exposure to synthetic and manipulated content leads audiences to distrust authentic reporting

alongside fabricated content, weakening the capacity of legitimate journalism and electoral communication to inform voters at all (Li, 2025). Neither of these effects requires a large number of successful fabrications; a small number of high-profile cases can be enough to shift the baseline level of public trust.

It is worth being precise about what these threats do and do not establish on their own. Detection and verification remain a persistent weakness. Forensic tools for identifying synthetic audio and video exist, but the pace of improvement in generative models tends to outstrip the pace of detection research, and Nigerian fact-checking organisations operate with far fewer resources than their counterparts in wealthier democracies (Loth et al., 2026). This is less a Nigeria-specific failure than a structural feature of the current technological moment, but its consequences fall more heavily on electoral systems with weaker verification infrastructure.

### **3.3 Nigerian Electoral Vulnerabilities**

AI capability alone explains relatively little about why disinformation has the effects it does; the effect depends on the terrain it lands on. Nigeria's political competition is structured around what Okezie et al. (2025) describe as a winner-takes-all logic, in which elections are contested less as a choice between policy platforms than as a struggle for ethnic and regional dominance. Disinformation that plays on ethnoreligious identity does not need to persuade a sceptical audience of something implausible; it needs only to confirm suspicions that many voters already hold about a rival region or faith community. This is the sense in which AI capability and Nigerian contextual vulnerability combine to produce electoral security risk: the same fabricated audio clip that might be treated with scepticism in a less polarised environment can instead function as confirmation of an existing narrative.

Institutional trust compounds the problem. Nigerians' confidence in the Independent National Electoral Commission was already fragile before 2023, and the well-documented technical failures of the Bimodal Voter Accreditation System and the IReV results portal, which was unable to upload polling unit results in real time as promised, gave disinformation about rigging a foothold in genuine institutional shortcomings rather than pure invention (Asangausung et al., 2025; Madueke & Enyiazazu, 2025). Disinformation is generally more persuasive when it can attach itself to a real grievance, and INEC's technological performance supplied exactly that kind of anchor. The 2023 election recorded a turnout of roughly 27 per cent, a marked decline from the 35 per cent recorded in 2019, and while turnout decline has multiple causes, including insecurity and logistical failures at polling units, several of the sources reviewed here treat the compounding effect of disinformation-fuelled distrust in the process as a contributing factor to voter disengagement (Namadi & Yerima, 2025; Akinyetun et al., 2021).

The information environment itself is a further vulnerability specific to the Nigerian and broader West African context. WhatsApp functions as a primary channel for political news and rumour, and its end-to-end encryption, while a legitimate privacy protection, also makes disinformation circulating within it largely invisible to platform moderators, fact-checkers, and researchers (Alberts, 2025). This closed-loop structure differs meaningfully from the largely open, publicly indexable platforms on which most disinformation research in the United States and Europe has been conducted, which is one reason findings from those contexts do not transfer neatly to Nigeria. Layered onto this is a media and digital literacy gap: survey-based assessments describe Nigerian digital literacy as low to moderate, which affects the public's capacity to distinguish synthetic from authentic content and increases susceptibility to disinformation that flatters existing political loyalties (Akinyetun et al., 2021; Shalgan & Obateru, 2025). Political information warfare in Nigeria has also become an increasingly professionalised undertaking, with parties and campaigns reportedly commissioning influencers to circulate manipulated content on a paid basis, which further blurs the line between organic political expression and coordinated manipulation.



The consequences of this interaction were not confined to information effects. The 2023 election cycle saw physical attacks on at least twenty INEC facilities, and while attributing any single act of violence to disinformation exposure is empirically difficult, the broader climate of suspicion generated by rigging narratives, amplified through both AI-assisted content and older forms of rumour, is treated in the literature as a contributing condition rather than an incidental backdrop (Asangausung et al., 2025).

### 3.4 Critical Synthesis of Threat Literature

Read as a whole, the literature supports a reasonably confident set of claims and a more cautious set of others. It is well established that generative AI has industrialised the production of disinformation, lowering the cost and skill threshold required to produce convincing synthetic content, and that this shift was visible in Nigeria's 2023 election through documented deepfake audio and video (Okezie et al., 2025; Loth et al., 2026). It is similarly well supported that visual and audio disinformation produces stronger belief and attitude effects than text-based falsehoods, a finding replicated across multiple national contexts, and that Nigeria's specific combination of ethnoreligious cleavage, institutional distrust, and closed messaging platforms creates conditions in which such content travels with unusual force (Dan, 2025; Alberts, 2025).

Plausible but less firmly demonstrated is the claim that AI-enabled disinformation has directly altered electoral outcomes or voting behaviour at scale in Nigeria. The literature documents correlation and plausible mechanism, disinformation exposure coinciding with declining turnout and rising distrust, but rarely offers the kind of individual-level behavioural data that would establish a direct causal chain from exposure to vote choice or abstention. Khan et al. (2026) are explicit on this point, characterising AI-generated disinformation as a risk multiplier that interacts with pre-existing identity threat and institutional distrust to raise the probability of unrest, rather than a sufficient cause of violence on its own. This is a meaningfully different claim from saying AI disinformation causes electoral violence, and the distinction matters for how the threat should be weighed against other contributing factors such as security force deployment, logistical failures, and long-standing communal tension.

Genuinely speculative, by contrast, is the prospect of a coming generation of multimodal systems capable of sustaining entirely fabricated virtual personas across text, image, video, and real-time interaction, sometimes referred to in the literature as a next-generation synthetic media threat. This scenario is technically plausible given the trajectory of current models, but it has not yet been observed in an actual electoral cycle anywhere, Nigeria included, and treating it as an imminent certainty would outrun the available evidence.

Two evidentiary limitations deserve emphasis before moving to the regulatory analysis in Section 4. First, there remains a shortage of Nigerian survey and experimental data on how ordinary voters actually perceive and respond to different forms of AI-generated content; most existing Nigerian scholarship relies on qualitative case documentation and content analysis of the disinformation itself rather than measurement of audience response. Second, the field as a whole is constrained by the opacity of platform data, since researchers depend on public content and self-reported exposure rather than the engagement and reach metrics held by the platforms themselves. These limitations do not undermine the case that AI-enabled disinformation constitutes a genuine and escalating threat to Nigeria's electoral security. They do mean that claims about the precise scale of its electoral impact should be stated with a degree of caution; the existing evidence base, however, is more than sufficient to justify the regulatory and institutional questions taken up next.

## 4. REGULATORY AND GOVERNANCE LANDSCAPE

### 4.1 Global and Nigerian Regulatory Framework

Governments have moved, over the past several years, from treating platform self-regulation as sufficient toward what several scholars describe as a model of descending co-regulation, in which binding legal obligations sit above voluntary industry codes rather than replacing them entirely. The European Union illustrates this shift most fully through what amounts to a regulatory triangle: the EU AI Act classifies AI systems by risk tier and imposes obligations accordingly, the Digital Services Act holds large platforms accountable for how their algorithms amplify content, and the European Media Freedom Act protects pluralism in the news environment those platforms distribute. The United States has taken a markedly different path, relying on executive orders and sector-specific rules rather than a comprehensive statute, which leaves considerably more discretion with individual agencies and platforms. Elsewhere on the continent, Kenya's National AI Strategy for 2025 to 2030 sets out an ethics-first approach to governance, while South Africa and Rwanda have built mandatory risk assessments directly into their sectoral regulatory frameworks rather than treating AI governance as a standalone concern.

Nigeria's regulatory posture reads as reactive against this backdrop, a characterisation Durodola (2024) makes explicitly when contrasting the country's approach with the more anticipatory models emerging elsewhere. The Electoral Act 2022 supplies the legal foundation for INEC's technological infrastructure, including the Bimodal Voter Accreditation System and the IReV portal, but it was drafted without any specific contemplation of generative AI and offers no guidance on synthetic media, automated content, or algorithmic manipulation during campaigns. The Nigeria Data Protection Act of 2023 and the Startup Act of the same year address data privacy and the innovation ecosystem respectively, and both matter for how AI systems handle Nigerian users' information, but neither establishes a risk classification scheme for AI applications of the kind now standard in the EU framework. The instrument most often invoked against online disinformation remains the Cybercrime (Prohibition, Prevention) Act of 2015, whose Section 24 criminalises the transmission of false information via computer systems. That section was written for a pre-generative-AI internet and has been repeatedly criticised for its use against journalists and government critics rather than against the kind of coordinated, AI-assisted manipulation this review is concerned with.

Two further legislative efforts sit in an unresolved state. The Protection from Internet Falsehood and Manipulation Bill of 2019 and the National Commission for the Prohibition of Hate Speech Bill have both drawn sustained opposition from civil society and media organisations, largely because their definitions of prohibited speech are broad enough to capture legitimate political criticism alongside genuine disinformation. Neither has become law, which leaves Nigeria without a dedicated disinformation statute even as the volume of AI-generated political content continues to grow.

Institutional responsibility is similarly dispersed. INEC itself established an AI Division in 2025, an acknowledgment that election management now requires some in-house technical capacity to address AI-related risks, though the division's mandate so far leans toward operational uses such as geospatial planning for materials distribution rather than disinformation detection specifically. Broadcasting and telecommunications oversight is split between the Nigerian Communications Commission, the National Broadcasting Commission, and the National Information Technology Development Agency, three bodies whose licensing, content-monitoring, and digital policy functions overlap in ways that were not designed with generative AI in mind. The National Centre for Artificial Intelligence and Robotics, created in 2020, was conceived primarily as a technical research institution rather than a regulatory one, and its work on AI capability building sits somewhat apart from the enforcement questions this section is concerned with.

## 4.2 Regulatory Gaps

The picture that emerges from this framework is not one of complete regulatory absence. Nigeria has cybercrime law, data protection law, and an electoral statute that at least contemplates digital voting infrastructure. The more accurate description of the problem is a mismatch between what these instruments were built to address and what generative AI now does. The Cybercrime Act's provisions on false information were written for a period when disinformation meant a rumour or a doctored image, not a synthetic video indistinguishable from authentic footage or an autonomous bot capable of sustaining a humanlike conversation over days. Nothing in current Nigerian law defines deepfakes, synthetic media, or agentic AI systems as distinct categories requiring distinct treatment, and no statute offers a definition of disinformation itself precise enough to survive serious constitutional challenge, a gap Kortukova and Mazarakis (2026) treat as characteristic of AI-era disinformation law more broadly rather than unique to Nigeria.

Institutional fragmentation compounds the definitional problem. Because the NCC, NBC, and NITDA hold overlapping mandates over digital content, responsibility for acting on a piece of AI-generated disinformation during an election can be genuinely unclear, and each agency has some basis for treating it as another agency's problem. Even where responsibility is clear, technical capacity often is not. The NBC, in particular, lacks the forensic tooling needed to verify whether a given piece of audio or video circulating during a campaign was synthetically produced, which means enforcement frequently depends on external fact-checking organisations rather than the regulator itself. This capacity gap sits inside a broader pattern that Kortukova and Mazarakis (2026) describe as a policy response lag, in which legislation is drafted to address threats that have already been superseded by the time it passes, a dynamic made worse by the availability of open-source generative models that carry no watermarking obligations at all.

The transnational character of the platforms carrying this content raises a separate set of problems that national legislation is poorly equipped to solve on its own. A Nigerian statute can regulate what happens within Nigerian jurisdiction, but the platforms distributing AI-generated political content, and frequently the AI models producing it, operate across borders that Nigerian regulators have no authority over. This produces what several scholars term a legal grey zone around liability: it remains unsettled whether responsibility for harmful synthetic content should rest with the company that built the underlying model, the platform that hosted and amplified it, or the individual who deployed it, and Nigerian law offers no clear answer.

Finally, any regulatory response has to contend with the tension between restricting harmful content and protecting legitimate political speech. The penalties proposed in some drafts of Nigeria's falsehood and hate speech bills have included capital punishment for the most serious offences, a severity that Inobemhe and Ossai (2026) argue creates strong incentives for journalists and ordinary citizens to self-censor rather than risk prosecution under vaguely worded provisions. Platforms face a parallel pressure in the opposite direction: accountability regimes that impose steep fines for hosting harmful content can push platforms toward removing borderline material indiscriminately rather than risk liability, producing the kind of over-censorship that chills legitimate debate as effectively as the disinformation it was meant to address. Nigeria's regulatory challenge, in other words, is not simply building enforcement capacity against AI-enabled disinformation, but doing so without recreating the censorship risks that have made previous legislative attempts so difficult to pass.

## 5. CRITICAL EVALUATION OF EXISTING POLICY RESPONSES

### 5.1 Existing Policy Responses

Government and electoral institutions have made some concrete moves beyond legislation. INEC's AI Division represents the most direct institutional response within Nigeria's electoral apparatus, using geospatial intelligence to plan the distribution of voting materials and, in principle, to support fraud detection, though its disinformation-specific capabilities remain limited relative to its logistical functions (Haruna, 2025). At the continental level, the African Union has proposed a Continental AI Strategy intended to harmonise governance approaches across member states, but the Malabo Convention on cybersecurity and data protection that would underpin much of this cooperation has been ratified by only fifteen countries, which leaves the AU's ambitions considerably ahead of what member states have actually committed to in binding form.

Technology platforms have introduced their own policies independent of any government mandate. Meta, X, and TikTok have each adopted labelling systems for manipulated media and deployed AI tools to flag hate speech and coordinated inauthentic behaviour. These are, in the language some scholars use, *sui generis* interventions, developed unilaterally by each platform rather than imposed by a common regulatory standard, and enforcement varies considerably both across platforms and across languages and regions within a single platform. Nigerian content, produced largely in English alongside Nigerian Pidgin and several indigenous languages, tends to receive less consistent moderation than content in major global languages, a pattern that recurs across the disinformation literature on the Global South.

Civil society has filled a substantial part of the gap left by weak state and platform enforcement. Dubawa in Nigeria, alongside Africa Check and Kenya's PesaCheck, function as what Roy et al. (2025) call retroactive gatekeepers, organisations that verify claims after they have already begun circulating rather than preventing their spread in the first place. Several of these organisations have moved toward a participatory model, inviting audience members to submit suspicious content for verification via WhatsApp, which has the practical advantage of meeting disinformation on the same closed platform where much of it originates. UNESCO's Global Media Literacy Framework, launched in 2025, represents an attempt to address the problem further upstream, and grassroots initiatives in several African countries have begun experimenting with pre-bunking, exposing audiences to weakened versions of manipulation techniques before they encounter the real thing, on the theory that inoculation works better than correction after the fact. News organisations covering hyper-local and multilingual disinformation increasingly rely on local stringers who understand the specific communities and languages involved, since automated tools generally cannot.

### 5.2 Critical Assessment of Effectiveness

Taken individually, several of these interventions show real effect. Hoes et al. (2024) find that prominent misinformation interventions, including fact-checking and platform labelling, do measurably reduce belief in the specific false claims they target. The complication is what happens beyond the targeted claim. The same interventions tend to produce a spillover effect, in which repeated exposure to correction and warning labels increases scepticism not just toward the debunked content but toward information in general, including accurate reporting from credible sources. A public that has learned to doubt everything is not obviously better protected than one that occasionally believes something false, and this trade-off runs through much of the literature on fact-checking effectiveness rather than being unique to the Nigerian case.

Technical interventions carry their own limitations. Watermarking, often proposed as a solution for

identifying AI-generated content at the point of creation, is not currently tamper-resistant; straightforward post-processing, cropping a video, recompressing an image, can strip watermarks entirely, which makes legal mandates requiring disclosure difficult to enforce in practice even where such mandates exist. The pace problem compounds this. Fact-checking and formal regulatory response both operate on a timescale of hours to days, while algorithmically amplified content can reach its peak audience within minutes, so even an accurate and well-resourced correction frequently arrives after the content it addresses has already done most of its persuasive work. Layered on top of both problems is a language barrier specific to contexts like Nigeria: most automated moderation systems and AI detection tools are trained predominantly on high-resource languages, leaving Nigerian Pidgin and indigenous languages considerably less well protected than English-language content on the same platforms, an asymmetry that mirrors the broader Global North bias identified earlier in this review's conceptual discussion.

Some of the more troubling findings concern effects that run in the opposite direction from what these interventions intend. The liar's dividend identified in earlier sections reappears here in policy terms: the more platforms and governments publicise the existence and sophistication of deepfakes, the easier it becomes for politicians to dismiss authentic, damaging recordings as fabrications, which means awareness campaigns can inadvertently strengthen the very evasion tactic they are meant to counter. A related and broader consequence is what several sources describe as a general trust crisis, in which the public stops being able to identify a reliable arbiter of truth at all, government, media, and platform alike having lost some claim to authority in this domain. This raises a genuine democratic trade-off in the policy responses currently favoured. Handing large technology companies expanded authority to label, remove, and adjudicate political content positions private firms as *de facto* arbiters of political truth, a role none of them were designed for and none are democratically accountable to perform, and the chilling effect this produces on ordinary political speech is not obviously smaller than the disinformation problem it is meant to solve.

## **6. RESEARCH GAPS AND FUTURE RESEARCH AGENDA**

### **6.1 Current Knowledge Gaps**

The preceding sections point toward a literature that is conceptually active but empirically thin, particularly where Nigeria is concerned. Several gaps stand out as genuine limits on what can currently be said with confidence.

The conceptual foundation itself carries an unresolved bias. As Section 2 established, most theorising about disinformation, and about AI's effect on it, originates in Global North contexts built around open, publicly indexable platforms. Nigeria's information environment runs substantially through closed, encrypted channels, and existing theory has not been reworked with this structural difference in mind. A related conceptual gap concerns the disproportionate scholarly attention paid to deepfakes relative to the broader category of AI-generated persuasive content that falls short of outright fabrication, work Beacken et al. (2026) describe as generative propaganda, which likely does more everyday electoral work than the more dramatic cases that dominate coverage.

The empirical gap is more acute still. Nigerian scholarship on AI-enabled disinformation is overwhelmingly qualitative, built from case documentation of specific incidents such as the 2023 deepfake endorsements and rigging audio, rather than from survey or experimental data capturing how ordinary voters actually perceive and respond to synthetic content. This leaves a basic question unanswered: does exposure to AI-generated disinformation change what Nigerian voters believe or how they vote, or does it mainly reinforce positions voters already hold. Section 3.4 was explicit that the literature supports correlation and plausible mechanism, disinformation exposure alongside declining turnout and rising



institutional distrust, without the individual-level data needed to establish a causal pathway. Khan et al.'s (2026) framing of AI disinformation as a risk multiplier rather than a sufficient cause captures the current state of evidence honestly, but it is a conclusion reached in the absence of the data that would settle the matter more precisely.

Methodologically, the field is constrained by platform data opacity. Researchers work from public content and self-reported exposure because the reach and engagement data that would allow more rigorous measurement sits inside platforms that do not make it available. This is not solvable by better research design alone; it requires a change in what data platforms are willing, or required, to release.

The literature is also almost entirely cross-sectional. Nearly everything written about Nigerian electoral disinformation concentrates on the 2023 cycle, which means there is no basis yet for judging whether the threat is escalating, stabilising, or provoking adaptive responses from voters and institutions over time. Comparative grounding is similarly thin: Nigeria is frequently cited alongside Kenya, South Africa, or India as evidence of a broader Global South pattern, but few studies systematically compare how similar AI capabilities interact with different institutional and media environments across African states.

Finally, Sections 2 and 4 both gesture toward AI, platforms, and institutions as an interacting system, institutionalism and complexity theory were adopted precisely because no single actor produces electoral security outcomes alone, yet very little empirical work actually models this interaction. Studies tend to isolate one node, a platform's amplification behaviour, or INEC's technical performance, or a piece of regulation, rather than tracing how weaknesses in one component compound weaknesses in another.

## 6.2 Future Research Priorities

Addressing the empirical gap should come first, since so much of the regulatory and theoretical discussion in this review rests on inference rather than direct measurement. Nigeria-specific survey and experimental research, testing how voters across different regions and literacy levels actually respond to synthetic audio, video, and text, would supply the evidence current scholarship lacks and would let researchers distinguish genuine persuasion effects from mere exposure.

Longitudinal research spanning multiple electoral cycles matters for a related reason: a single election cannot show whether AI-enabled disinformation is a passing shock that institutions and voters adapt to, or a structural feature of Nigerian elections going forward. Comparative African research would help separate what is specific to Nigeria from what reflects a broader regional pattern, informing which regulatory lessons are transferable from Kenya or South Africa and which are not.

On the technical side, detection and verification research needs to move beyond tools built for high-resource languages. Given the language barrier identified in Section 5.2, research on AI detection adapted to Nigerian Pidgin and indigenous languages is not a marginal concern but a precondition for equitable protection. Platform and algorithmic governance research, particularly work examining what data access would let independent researchers evaluate amplification patterns, would address the opacity problem identified above rather than working around it.

Regulatory effectiveness research is conspicuously absent from the current literature. Nigeria has cybercrime law, data protection law, and an evolving institutional apparatus, but no study yet asks whether any of these instruments actually reduces the incidence or impact of AI-enabled disinformation. Pairing this with institutional capacity research, assessing what the NBC, NCC, and NITDA would need to detect synthetic content at scale, would connect legal analysis to operational reality.

None of these priorities can be pursued adequately within a single discipline. The interaction between technological capability, political behaviour, institutional performance, and legal structure that this review

has traced calls for interdisciplinary work combining computer science, political science, law, and communication research, since each discipline working alone captures only one node of a problem that is, by its nature, systemic.

## 7. CONCLUSION

This review set out from a simple observation: AI has changed the production of disinformation more than it has changed the motives behind it. Generative AI is not a new political actor so much as a force multiplier for propaganda techniques that predate it by decades, adding scale, speed, and a realism that makes fabricated audio and video difficult to distinguish from the genuine article. That distinction, established in Section 2 and traced through the Nigerian evidence in Section 3, has organised the argument throughout. AI-enabled disinformation matters for electoral security because it industrialises the exploitation of grievances that already exist, not because it invents new ones. Its most persuasive forms, synthetic video and audio, carry an evidentiary weight that text-based rumour never had.

The relationship to electoral security runs through several connected mechanisms. Synthetic content erodes public trust in electoral institutions when it targets bodies like INEC directly. It produces a liar's dividend that lets political actors dismiss genuine evidence as fabricated, a dynamic Nigeria's own 2023 election already demonstrated. And it compounds the kind of epistemic fatigue in which audiences stop trusting information generally, verified and fabricated alike, rather than causing that fatigue on its own. AI did not invent Nigeria's ethno-religious cleavages or its winner-takes-all political culture; it supplied a more efficient means of activating divisions that were already there.

This is why the Nigerian context did more analytical work in this review than a simple case study role. Nigeria's closed, encrypted information environment, its uneven digital literacy, and the well-documented technical failures of the BVAS and IReV systems during the 2023 election gave AI-generated disinformation genuine institutional grievances to attach itself to, rather than requiring it to invent suspicion from nothing. Complexity theory and institutionalism, adopted in Section 2 as complementary lenses, proved useful precisely because no single factor, technology, institutions, or political culture, explains the 2023 electoral security failures alone. It was the interaction among all three that mattered.

The regulatory analysis in Sections 4 and 5 reached a correspondingly qualified conclusion. Nigeria's problem is less an absence of law than a mismatch: existing instruments were drafted before generative AI was a serious electoral concern, and they are now being asked to address a threat they were never built for. Institutional fragmentation across the NCC, NBC, and NITDA, weak technical capacity for verifying synthetic media, and the unresolved tension between restricting harmful content and protecting legitimate political speech together explain why Nigeria's response has remained reactive rather than anticipatory. The interventions currently in place, from INEC's AI Division to Dubawa's fact-checking work, show real but partial effectiveness, constrained by the same structural features, platform opacity, language barriers, and the speed of algorithmic amplification, that make the underlying threat difficult to counter in the first place.

The contribution of this review lies less in identifying a threat that was previously unknown than in synthesising a fragmented literature into a single analytical account of how AI capability, Nigerian institutional and political conditions, and regulatory capacity interact to produce electoral security risk. Existing scholarship has documented pieces of this picture, deepfakes in the 2023 election, INEC's technical shortcomings, the weaknesses of Nigeria's cybercrime statute, without consistently connecting them into the systemic account Sections 2 through 5 have attempted here. Whether that account holds depends on the empirical work identified in Section 6, particularly the Nigeria-specific evidence on actual voter response that the field currently lacks. For now, Nigeria's electoral security depends on institutional

and regulatory capacity keeping pace with a technology whose trajectory shows no sign of slowing. That gap, more than any single deepfake or piece of legislation, is the problem this review has tried to make legible.

## References

- Akinyetun, T. S., Odeyemi, D. D., & Alausa, J. A. (2021). Social media and electoral violence in Nigeria: Sustainable Development Goal 16, a panacea? *KIU Interdisciplinary Journal of Humanities and Social Sciences*, 2(2), 169–194.
- Alberts, K. (2025). Unravelling the web: Exploring the dynamics of online disinformation and its impact on democracy in Africa. *Canadian Journal of African Studies / Revue canadienne des études africaines*, 59(2), 333–359. <https://doi.org/10.1080/00083968.2025.2491679>
- Asangausung, O., Udoms, B., & Yağcı, A. (2025). Managing election-related violence in Nigeria: Focus on the use of technologies in the 2023 General Elections. *Yıldız Social Science Review*, 11(1), 44–60.
- Beacken, G. D., Trauthig, I. K., & Woolley, S. C. (2026). Generative AI, propaganda, and digital authoritarianism: Comparative insights from six democratically weakened countries. *Big Data & Society*, April–June, 1–14. <https://doi.org/10.1177/20539517251410064>
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. <https://doi.org/10.1177/0267323118760317>
- Dan, V. (2025). Deepfakes as a democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. *The International Journal of Press/Politics*, 31(3), 525–550. <https://doi.org/10.1177/19401612251317766>
- Doshi, J., Novacic, I., Fletcher, C., Borges, M., Zhong, E., Marino, M. C., Gan, J., Mager, S., Sprague, D., & Xia, M. (2024). *Sleeper social bots: A new generation of AI disinformation bots are already a political threat*. arXiv. <https://arxiv.org/abs/2408.12603>
- Durodola, O. (2024). *Bridging the AI governance gap: Ethical and regulatory imperatives for generative AI in Nigeria*. Babcock University.
- Ekpang, J. E., Iyorza, S., & Ekpang, P. O. (2023). Social media and artificial intelligence: Perspectives on deepfakes' use in Nigeria's 2023 general elections. *KIU Interdisciplinary Journal of Humanities and Social Sciences*, 4(2), 109–124. <https://doi.org/10.59568/KIJHUS-2023-4-2-08>
- Ferrara, E. (2025). Charting the landscape of nefarious uses of generative artificial intelligence for online election interference. *First Monday*, 30(6). <https://dx.doi.org/10.5210/fm.v30i6.14188>
- Fisher, S. A., Howard, J. W., & Kira, B. (2024). Moderating synthetic content: The challenge of generative AI. *Philosophy & Technology*, 37, Article 133. <https://doi.org/10.1007/s13347-024-00818-9>
- Haruna, U. D. (2025). Strengthening the Nigerian electoral process through artificial intelligence adoption: A review. *International Journal of Development Mathematics*, 2(4), 254–261.
- Hassan, I. (2023). Dissemination of disinformation on political and electoral processes in Nigeria: An exploratory study. *Cogent Arts & Humanities*, 10, Article 2216983. <https://doi.org/10.1080/23311983.2023.2216983>

- Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour*, 8, 1545–1553. <https://doi.org/10.1038/s41562-024-01884-x>
- Inobemhe, K., & Ossai, C. O. (2026). Media regulation in Nigeria's digital age: A systematic review of policy frameworks. *Academia Journal of Humanities & Social Sciences*, 3, 156–175.
- Islas, O., Gutiérrez, F., & Arribas, A. (2024). Artificial Intelligence, a powerful battering ram in the disinformation industry. *New Explorations*.
- Khan, F., Khan, T. Z., & Malik, N. (2026). Artificial Intelligence-driven disinformation campaigns and their influence on political violence and extremism. *Social Sciences Spectrum*, 5(3), 357–374. <https://doi.org/10.71085/sss.05.03.586>
- Kortukova, T., & Mazaraki, A. (2026). Legal regulation of countering disinformation in the age of AI. *AGORA International Journal of Juridical Sciences*, 20(1), 146–154.
- Li, S. (2025). The social harms of AI-generated fake news: Addressing deepfake and AI political manipulation. *Digital Society & Virtual Governance*, 1(1), 72–88. <https://doi.org/10.6914/dsvg.010105>
- Loth, A., Kappes, M., & Pahl, M. O. (2026). The verification crisis: Expert perceptions of GenAI disinformation and the case for reproducible provenance. *Companion Proceedings of the ACM Web Conference 2026 (WWW Companion '26)*. <https://doi.org/10.1145/3774905.3795484>
- Madueke, K. L., & Enyiazu, C. (2025). Electoral integrity and election management in Nigeria: The case of the 2023 General Election. *World Affairs*, 1–15. <https://doi.org/10.1177/00438200251205510>
- Mansur, M. A. (2025). AI and cyber-enabled threats to democracy through algorithmic manipulation and generative AI in undermining democratic integrity. *ESI Preprints*. <https://doi.org/10.19044/esipreprint.8.2025.p658>
- Namadi, M. M., & Yerima, H. (2025). An assessment of the impact of election observation and public participation on the democratic process in Nigeria: A case study of Kano State. *International Journal of Law, Politics & Humanities Research*, 8(6), 281–298.
- Okezie, W. I., Inobemhe, K., Udeh, N. S., & Modeyin, O. E. (2025). AI in Nigerian elections: Tool for safeguarding democracy or weapon for deepfakes and disinformation. *Nepal Journal of Multidisciplinary Research (NJMR)*, 8(3), 89–108. <https://doi.org/10.3126/njmr.v8i3.79017>
- Okolo, C. T. (2024). *African democracy in the era of generative disinformation: Challenges and countermeasures against AI-generated propaganda*. Brookings Institution / Research ICT Africa.
- Olawunmi, K. (2025). Computational propaganda, disinformation, and democracy: Multidisciplinary strategies for 2025. *International Journal of Social Science and Human Research*, 8(9), 6874–6880. <https://doi.org/10.47191/ijsshr/v8-i9-19>
- Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society*, 1, Article 19. <https://doi.org/10.1007/s44206-022-00010-6>
- Romanishyn, A., Malytska, O., & Goncharuk, V. (2025). AI-driven disinformation: Policy recommendations for democratic resilience. *Frontiers in Artificial Intelligence*, 8, Article 1569115. <https://doi.org/10.3389/frai.2025.1569115>

- Roy, E., Laskar, K. A., & Reyaz, M. (2025). Retroactive gatekeepers: Fact-checkers, electoral disinformation, and audience engagement in India. *Emerging Media*, 3(4), 582–600. <https://doi.org/10.1177/27523543251395190>
- Shalgan, L. M., & Obateru, T. (2025). Media literacy, misinformation and disinformation on social media during the 2023 elections in Plateau State. *International Journal of Social Science Research and Anthropology (TIJSSRA)*, 7(6), 175–182.
- Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. *arXiv*. <https://arxiv.org/abs/2311.17394>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Vasist, P. N., Chatterjee, D., & Krishnan, S. (2023). The polarizing impact of political disinformation and hate speech: A cross-country configural narrative. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-023-10390-w>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>)