

ARTIFICIAL INTELLIGENCE AND THE CRISIS OF TRUST: MIS/DIS/MALINFORMATION IN JOURNALISM AND DEMOCRATIC COMMUNICATION IN NIGERIA

ALIYU HUSSAINI

Department of Journalism and Media Studies
Nuhu Bamalli Polytechnic, Zaria
08028432950 | aliyuh3@gmail.com

HAUWA'U ABUBAKAR BUHARI, Ph.D.

Department of General Studies, Nuhu Bamalli Polytechnic, Zaria
08033190975 | Jiddaturrahman2424@gmail.com

&

HUSSAIN ALHASSAN

Department of Strategic Communication
School of Communication and Information Technology
Nuhu Bamalli Polytechnic, Zaria
08038735373 | alhassanhussain2@gmail.com

Abstract

Democracy in Nigeria, as in many other nations, rests heavily on access to credible information and public trust in media institutions. The growing use of artificial intelligence (AI) technologies has introduced fresh strains on this relationship. Tools such as deepfakes, chatbots, and algorithmic news feeds now make it easier to create and circulate misinformation, disinformation, and malinformation (MDM) at speed and scale. This paper examines how AI-driven content is affecting journalism and democratic communication in Nigeria. Drawing on the 2019 and 2023 general elections, the COVID-19 pandemic, and narratives around ethnic and religious conflict, it considers both the ways AI amplifies harmful content and the limited but important roles it can play in countering it. The discussion assesses effects on media credibility, public trust, and the quality of democratic deliberation, and ends with practical suggestions for restoring trust through greater transparency, stronger media literacy, clearer regulation, and closer cooperation among journalists, technology developers, and government.

Keywords: Artificial Intelligence (AI), Misinformation, Disinformation, Media Trust, Democratic Communication

Introduction

Nigeria is Africa's largest democracy and its most populous country. Its media landscape is shaped by political competition, ethnic diversity, and religious difference. The press has long been expected to inform citizens, frame public debate, and support democratic processes. At the same time, the system has been vulnerable to political interference, bias, and, more recently, the disruptive effects of artificial intelligence. In Nigerian newsrooms and on social platforms, AI now works in two directions. On one side, it can speed up production, support data analysis, and help fact-checkers. On the other, the same technologies deepfakes, automated text generation, and recommendation algorithms make it cheaper and faster to produce and spread misleading material. In a society already marked by political and communal tension, these capabilities raise the stakes of information disorder (Wardle & Derakhshan, 2017; Udoh, 2025).

Political and religious actors have begun to use AI to push false narratives, deepen polarisation, and shape electoral outcomes. Platforms such as Facebook, X (formerly Twitter), and WhatsApp remain central to political conversation. Algorithmic amplification means that both accurate and harmful content can travel far and fast. In an environment where trust in traditional media is already uneven, AI adds another layer of uncertainty to how Nigerians receive and judge political information (Shaibume & Avindigh, 2026; Okocha, 2026).

Statement of the Problem

The arrival of AI tools in Nigeria's media space has created clear opportunities and equally clear risks. AI can assist news production and fact-checking, yet it has also intensified the crisis of trust. Realistic but fabricated audio, video, and text make it harder for citizens to separate genuine reporting from manufactured content. During the 2019 and 2023 elections, bots and deepfake videos were used to circulate false claims about candidates and results (Udoh, 2025; Shaibume & Avindigh, 2026).

The central difficulty is that AI blurs the boundary between fact and fiction. In a media system already strained by bias and political pressure, this technology amplifies existing problems and contributes to the wider circulation of misinformation, disinformation, and malinformation. As a result, public confidence in both legacy and digital media continues to weaken, and the quality of democratic communication suffers (Okocha, 2026; Mahmud, Ogunbodede & Ewata, 2025).

It is against this backdrop that this study decided to examine how AI-driven content is affecting journalism and democratic communication in Nigeria. Drawing on the 2019 and 2023 general elections, the COVID-19 pandemic, and narratives around ethnic and religious conflict, it considers both the ways AI amplifies harmful content and the limited but important roles it can play in countering it.

Objective of the Study

This paper examines the role of AI-driven misinformation, disinformation, and malinformation in the Nigerian media landscape. It considers how these technologies are used to shape public perception, influence political outcomes, and erode trust in journalism. Through case material drawn from the 2019 and 2023 elections, COVID -19- related claims, and ethno -religious conflict narratives, the paper assesses effects on journalistic integrity and democratic communication. It also outlines practical responses ethical guidelines, media literacy efforts, and regulatory measures that may help rebuild trust.

Research Questions

- a. How do algorithmic news feeds and AI-generated content influence perceptions of credibility and trust in Nigerian media?
- b. What concrete examples show AI-driven misinformation or disinformation damaging journalistic credibility or public trust in Nigeria?
- c. What technological, ethical, and regulatory strategies appear most

promising for protecting or restoring trust in journalism in the Nigerian AI era?

Review of Related Literature

Misinformation, Disinformation, and Malinformation

Scholars distinguish three related forms of information disorder. Misinformation is false content shared without intent to harm. Disinformation is deliberately fabricated material designed to deceive. Malinformation involves the strategic use of accurate information often private or decontextualised to damage individuals or groups (Wardle & Derakhshan, 2017).

In Nigeria, misinformation travels quickly on WhatsApp and Facebook. During the COVID-19 pandemic, unfounded claims about cures and prevention measures circulated widely and complicated public-health messaging (BBC Africa Eye, 2020). Disinformation campaigns have been more organised, especially around elections. In 2023, AI-assisted bots helped distribute manipulated images, doctored videos, and false statements about candidates and results, contributing to public doubt about the process (Udoh, 2025; Ibrahim, 2023). Malinformation has also appeared in the form of private recordings released at politically sensitive moments. In a fragile democracy, such tactics raise the cost of political competition and weaken confidence in institutions (Shaibume & Avindigh, 2026).

AI and Media in Nigeria

AI occupies a dual position in Nigerian journalism. Fact-checking organisations such as Africa Check and Dubawa use automated tools to verify claims during elections and major news events (Dubawa, 2023; Ajakaiye, cited in Ndhlovu, 2023). At the same time, the same technologies enable highly convincing synthetic content that can be distributed at scale on X and Facebook (Okocha, 2026). Newsrooms including Premium Times and The Cable have begun experimenting with automated reporting, data-driven journalism, and content personalisation. These applications can increase

speed and volume, yet they also raise questions about human oversight, the reproduction of bias, and the loss of contextual judgement. When AI-generated or AI-amplified material touches sensitive political or communal issues, the potential for harm grows (Udoh, 2025).

Algorithmic Personalisation, Feeds, and Amplification

Platforms popular in Nigeria rely on recommendation systems that prioritise engagement. The resulting filter bubbles limit exposure to opposing views and can intensify polarisation (Pariser, 2011; Bakshy, Messing & Adamic, 2015). During election periods the effect is especially pronounced: partisan and often misleading content travels further while corrective information struggles to keep pace (Udoh, 2025).

The 2023 elections illustrated the problem. AI-driven bots and coordinated networks amplified false claims about candidates, electoral violence, and results. Deepfake material further complicated verification. Because information access remains uneven across Nigeria, algorithmic amplification can widen existing divides and make democratic deliberation more difficult (Shaibume & Avindigh, 2026; Centre for Democracy and Development, 2023).

Trust, Credibility, and Public Perceptions

Survey evidence continues to show higher trust in radio and television than in social media among Nigerian audiences (Reuters Institute, 2023, 2024, 2025). Yet the growing presence of AI-mediated content makes source evaluation harder. Journalists themselves have voiced concern that AI may accelerate the spread of disinformation and weaken the perceived authority of established outlets (Okocha, 2026; Mahmud et al., 2025). Maintaining trust will therefore require transparent practices, clear labelling of synthetic content, and sustained human editorial control.

Theoretical Framework

This study draws on three complementary perspectives to make sense of AI, information disorder, and trust in Nigerian journalism. The paper relies

on the theory of information disorder developed by Wardle and Derakhshan (2017). Their distinction among misinformation, disinformation, and malinformation provides a precise vocabulary for the forms of harmful content that circulate in Nigeria. The framework is especially useful because it focuses not only on the content itself but also on the agents who produce and share it and the motivations that drive them whether negligence, political gain, or deliberate harm.

Second, the analysis is guided by media systems dependency theory (Ball-Rokeach & DeFleur, 1976). In periods of social uncertainty elections, public-health crises, or communal conflict citizens become more dependent on media for orientation. When those media are themselves compromised by synthetic or algorithmically amplified falsehoods, the dependency relationship can deepen public anxiety and further erode institutional trust. Nigeria's recurrent political and security crises make this dynamic particularly relevant.

Third, the paper incorporates insights from the concept of algorithmic amplification and filter bubbles (Pariser, 2011; Bakshy et al., 2015). Recommendation systems that prioritise engagement tend to reinforce existing beliefs and limit exposure to corrective information. In a multi-ethnic and multi-religious society, these technical features can accelerate polarisation and reduce the common informational ground required for democratic deliberation. Together, these perspectives frame AI not merely as a technical novelty but as a force that reshapes the production, circulation, and reception of political information. They also supply criteria for evaluating interventions: measures that increase transparency, reduce dependency on unverified sources, and interrupt harmful amplification are theoretically better placed to protect trust.

Methodology

The study adopts a mixed-methods design that combines quantitative and qualitative approaches. The quantitative component measures public

exposure to AI-related content, levels of trust in different media, and awareness of specific AI tools. The qualitative component explores how journalists, editors, fact-checkers, and technology practitioners experience and evaluate AI in their daily work. Combining the two strands allows both breadth of public perception and depth of professional insight.

Sampling

For the survey, a stratified random sample of 300 respondents was drawn from urban and rural areas, different age groups, genders, and educational backgrounds. Participants were selected from lists of registered voters and active social-media users so that the sample would reflect a cross-section of the Nigerian population rather than any single demographic niche. For the qualitative phase, 30 key informants were chosen purposively: 15 journalists working in print, broadcast, and digital outlets; five media editors with experience of AI tools; five fact-checkers from organisations such as Africa Check and Dubawa; and five technology specialists familiar with AI applications in media.

Data Collection Instruments

The survey questionnaire measured exposure to AI-generated content, trust in legacy media, social media, and AI-assisted outlets, awareness of deepfakes and bots, and attitudes toward the ethical use of AI in journalism. Items included Likert-scale, multiple-choice, and open-ended questions. A pilot test was conducted to refine wording and sequence. Semi-structured interviews explored the practical use of AI in newsrooms, ethical concerns, experiences with election-related and crisis-related misinformation, and suggestions for policy or professional reform. Interviews were recorded (with consent), transcribed, and analysed thematically.

Case Study Analysis

Three cases were selected for closer examination because each involved documented instances of AI-assisted or algorithmically amplified false

content with measurable effects on public discourse:

1. The 2023 Nigerian presidential election, focusing on bots, deepfakes, and coordinated false narratives about candidates and results.
2. COVID-19 misinformation, examining how recommendation systems amplified unverified health claims and complicated official messaging.
3. Ethno-religious conflict narratives in parts of northern Nigeria, looking at the circulation of manipulated images and videos that inflamed communal tension.

Case Examples and Findings

2023 Nigerian Elections

During the 2023 presidential contest, AI-assisted tools contributed to a noisy and often deceptive information environment. Bots distributed doctored images and videos; deepfake techniques were used to create false statements attributed to candidates. Many citizens expressed doubt about both the integrity of the process and the reliability of the outlets reporting it. Udoh (2025) documents how fabricated voice clips and deepfake videos of leading candidates gained millions of views before being debunked by fact-checkers such as Dubawa and AFP. Shaibume and Avindigh (2026) similarly report that politically motivated actors exploited AI tools to distort facts and erode trust in electoral institutions. Ibrahim (2023) identifies multiple typologies of information disorder including disinformation and malinformation deployed in official campaign communications. These findings align with the Centre for Democracy and Development's (2023) observation that synthetic and manipulated media were used at an unprecedented scale.

COVID-19 Pandemic

False claims about miracle cures and vaccine risks spread rapidly on

platforms whose algorithms favoured engagement over verification. Fact-checkers and the Nigeria Centre for Disease Control struggled to keep pace. The episode demonstrated that AI-driven amplification can outrun correction and leave lasting residual distrust of official health communication (BBC Africa Eye, 2020). This pattern is consistent with broader observations that recommendation systems prioritise emotionally charged content, thereby accelerating the circulation of health-related misinformation (Wardle & Derakhshan, 2017).

Ethno-Religious Conflicts

In states such as Plateau and Kaduna, manipulated videos sometimes footage from distant conflicts re-labelled as local events circulated on social media. The difficulty of rapid visual verification allowed malicious actors to inflame existing communal grievances and, in some instances, to escalate violence. Mahmud, Ogunbodede and Ewata (2025) show how deepfakes exploit visual evidence bias and transform social capital into misinformation transmission channels, generating persistent reputational and social damage even after exposure the so-called "liar's dividend" phenomenon. Okocha (2026) further notes that such synthetic content exacerbates societal divisions and undermines the broader ecosystem of trust essential to democratic media.

Data Presentation

This section presents the main survey and interview findings and relates them to the research questions and the theoretical framework outlined earlier.

Quantitative Results

Table 1: Participant Demographics (n = 300)

Category	Frequency (n=300)	Percentage (%)
Gender		
Male	150	50.0
Female	150	50.0
Age Group		
18–24	100	33.3
25–34	90	30.0
35–44	70	23.3
45+	40	13.3
Location		
Urban	180	60.0
Rural	120	40.0
Education		
Secondary	50	16.7
Tertiary	250	83.3

Field survey 2026

Table 1 shows a balanced gender split, a relatively young sample (almost two-thirds under 35), a majority of urban respondents, and a high proportion of tertiary-educated participants. The profile is therefore weighted toward digitally connected citizens while still including rural voices.

Table 2: Perceptions of Trust in Media Platforms (n = 300)

Media Platform	Low Trust (%)	Moderate Trust (%)	High Trust (%)
Legacy media (TV, radio)	20	30	50
Social media (Facebook, X)	60	25	15
AI-assisted / fact-checking outlets	20	50	30

Field survey 2026

Table 2 indicates that legacy media retain the highest levels of trust, while social media platforms are viewed with considerable scepticism. AI-assisted fact-checking services occupy a middle ground: most respondents express moderate rather than high or low confidence. The pattern is consistent with Reuters Institute Digital News Report findings for Nigeria, which repeatedly show higher trust in traditional and established independent brands than in social media (Reuters Institute, 2023, 2024, 2025). It also aligns with media-systems-dependency expectations that audiences turn to

established sources in uncertain times (Ball-Rokeach & DeFleur, 1976).

Table 3: Awareness of Selected AI Tools (n = 300)

AI Tool / Phenomenon	Aware (%)	Unaware (%)
Deepfakes	75	25
AI bots used in election campaigns	55	45
AI-generated fake news images	80	20

Field survey 2026

Awareness of deepfakes and synthetic images is high, yet fewer respondents recognise the role of AI bots in election campaigns. The gap points to a continuing need for public education that goes beyond dramatic visual examples to the quieter, automated amplification of political messages. This finding resonates with Mahmud et al. (2025), who report severe trust erosion in Nigeria's digital information ecosystem and low institutional trust on messaging platforms despite high interpersonal trust.

Table 4: Key Themes from Interviews with Media Professionals (n = 30)

Key Theme	Frequency	Percentage (%)
AI as a tool for efficiency	20	66.7
Ethical concerns about AI in journalism	25	83.3
AI's role in amplifying misinformation	15	50.0
Need for clearer regulation	10	33.3

Field survey 2026

Interview participants largely accept that AI can improve speed and efficiency. At the same time, ethical unease is widespread. Half the informants directly linked AI to the amplification of misinformation, and one-third called for stronger regulatory attention. These professional views align with Okocha's (2026) qualitative interviews with Nigerian journalists, which found that deepfakes pose a significant threat to journalists' credibility and public trust, and with Shaibume and Avindigh's (2026) conclusion that institutional responses remain largely reactive and constrained by limited technical capacity.

Discussion of Findings

The findings speak directly to the three research questions and are supported by a growing body of scholarly work on AI, information disorder, and media trust in Nigeria and comparable contexts. First, algorithmic feeds and AI-generated content are associated with lower perceived credibility,

especially on social platforms. Legacy media still enjoy relative advantage, consistent with media-systems-dependency expectations that audiences turn to established sources in uncertain times (Ball-Rokeach & DeFleur, 1976). This pattern is corroborated by successive Reuters Institute Digital News Reports for Nigeria, which show persistently higher trust in traditional and independent brands than in social media (Reuters Institute, 2023, 2024, 2025). Mahmud et al. (2025) further document that only a minority of respondents express trust in social-media information, while interpersonal trust remains higher creating vulnerabilities when deepfakes circulate through trusted personal networks.

Second, the 2023 election, COVID-19, and conflict cases supply concrete instances in which AI-assisted or algorithmically amplified falsehoods damaged trust in both news outlets and political institutions. Udoh (2025) provides detailed evidence of deepfake videos and AI-generated audio that reached large audiences before debunking. Shaibume and Avindigh (2026) argue that such content shaped public perception and eroded confidence in electoral institutions. Ibrahim (2023) demonstrates that multiple typologies of information disorder were operationalised in campaign communications. These findings are consistent with Wardle and Derakhshan's (2017) framework, which emphasises the roles of agents, messages, and interpreters in the production and reception of polluted information. Okocha (2026) adds that the psychological toll on journalists anxiety, fear, and professional vulnerability further weakens the media's capacity to serve as a trusted intermediary.

Third, journalists themselves identify transparency, human oversight, and targeted regulation as necessary responses recommendations that map onto the theoretical emphasis on reducing harmful amplification and restoring verifiable informational common ground. The dual nature of AI is widely acknowledged: while it offers efficiency gains (recognised by two-thirds of interviewees), its capacity to amplify misinformation and erode

credibility is viewed as a serious threat (Okocha, 2026; Udoh, 2025). Taken together, the data suggest that AI is neither an automatic threat nor an automatic solution. Its effects depend on the incentives of platforms, the capacity of newsrooms, the literacy of audiences, and the presence or absence of enforceable rules. Without deliberate intervention, the balance currently tilts toward further erosion of trust. This conclusion is supported by both the empirical patterns reported here and the broader scholarly literature on synthetic media and democratic communication in Nigeria.

Policy Implications and Recommendations

Drawing on the empirical findings and the supporting literature, the following recommendations are offered for key stakeholders.

For media houses

- Label AI-generated or heavily AI-assisted content clearly so audiences can evaluate it (Okocha, 2026).
- Adopt internal ethics codes that keep human editors responsible for final publication decisions.
- Invest in continuous training so journalists understand both the capabilities and the limits of the tools they use.

For technology platforms

- Develop provenance and watermarking systems that make synthetic media easier to detect (Udoh, 2025).
- Adjust recommendation algorithms during high-stakes periods (elections, public-health emergencies) to prioritise verified sources.
- Maintain sustained partnerships with independent Nigerian fact-checking organisations such as Dubawa and Africa Check.

For government and regulators

- Craft transparent rules on AI disclosure and accountability without stifling legitimate journalistic innovation (Shaibume & Avindigh, 2026).
- Strengthen legal tools against coordinated AI-driven disinformation

campaigns, particularly around elections.

- Support public media-literacy programmes, especially in rural and low-literacy communities.

For civil society

- Expand community-level media literacy that teaches practical verification skills (Mahmud et al., 2025).
- Encourage citizens to engage with independent fact-checking services and to demand transparency from both platforms and newsrooms.

Conclusion

Artificial intelligence has already altered the production and circulation of news in Nigeria. It offers efficiencies and new storytelling possibilities, yet it also multiplies the scale and speed at which false or harmful content can travel. The resulting pressure on media credibility and public trust is visible in election periods, public-health crises, and moments of communal tension. Scholarly work by Udoh (2025), Shaibume and Avindigh (2026), Okocha (2026), Mahmud et al. (2025), and others confirms that these dynamics are not isolated anecdotes but part of a broader pattern of information disorder amplified by AI.

No single actor can reverse the damage. Journalists, news organisations, technology companies, regulators, and ordinary citizens all have roles to play. Transparent AI practices, enforceable rules, and sustained investment in media literacy form the most realistic path toward protecting the informational foundations of Nigerian democracy. If these measures are pursued with seriousness, AI need not remain a permanent threat to trust; it can become one more tool in the service of accurate, accountable public communication.

References

- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Ball-Rokeach, S. J., & DeFleur, M. L. (1976). A dependency model of mass-media effects. *Communication Research*, 3(1), 3–21.
- BBC Africa Eye. (2020). Nigeria's WhatsApp propaganda machine [Documentary]. BBC.
- Centre for Democracy and Development. (2023). Distorting Nigeria's elections? How disinformation was deployed in 2023. CDD.
- Dubawa. (2023). Fact-checking Nigeria's elections: Insights from the 2023 polls. Abuja: Dubawa.
- Ibrahim, A. A. (2023). A broth of ballots and falsehoods: A study of information disorder in the 2023 Nigerian presidential elections [Master's thesis, University of Iowa].
<https://doi.org/10.25820/etd.006859>
- Mahmud, M. O., Ogunbodede, K. F., & Ewata, T. O. (2025). Deepfakes as multimodal semiotic signs: Iconic, indexical, and symbolic dimensions in Nigerian social media. *Journal of Language and Linguistics of Communication*, 5(1), 98–120.
- Ndhlovu, L. (2023). Combating disinformation around Nigeria's 2023 election. International Journalists' Network.
- Okocha, M. U. (2026). Journalism in the shadows: AI deepfakes, credibility, and the fragility of media trust. *Journalism Practice*.
<https://doi.org/10.1080/17512786.2026.2719552>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Reuters Institute. (2023). Digital news report – Nigeria chapter. University of Oxford.
- Reuters Institute. (2024). Digital news report – Nigeria chapter. University of Oxford.
- Reuters Institute. (2025). Digital news report – Nigeria chapter. University of Oxford.
- Shaibume, B., & Avindigh, S. (2026). Artificial intelligence and the crisis of electoral integrity in Nigeria: Emerging threats of disinformation.

- Kashere Journal of Politics and International Relations, 4(1), 304–320.
- Udoh, I. U. (2025). AI and deepfakes on social media: A theoretical analysis of the technological disruption in Nigeria's 2023 general elections. Zenodo. <https://doi.org/10.5281/zenodo.18818547>
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe..