Published by

British Journal of Contemporary Research

BEXFORD PUBLISHING

Social Media Disinformation and National Security in Nigeria: A Critical Review of Digital Information Threats, Crisis Communication and Security Governance

Review Article | Volume 1 | Issue 6 | 2026 | Article Number: 011

Accepted: 13 October 2026 | Published: 10 November 2026 | ISSN: 2979-8582 (Online)

Crossref : **https://doi.org/10.67693/BJCR-T9PILQ3L**

Sakeena Audu

Political Science, Federal University Lafia, Nigeria

Correspondence: Sakeena Audu, Sakeena.audu2020@nda.edu.ng

Abstract

Nigerian public debate increasingly treats false content on social media as a matter of national security. This review asks how far the evidence supports that framing. It draws on political communication, computational social science, crisis communication and security studies, and follows the problem from the spread of disinformation through public trust and crisis communication to the state's governance response. Amplification is the best-documented link: false content travels faster than accurate content, and in Nigeria it gains authority when radio presenters and partisan commentators repeat it. Evidence that disinformation causes violence is thinner. Nigerian accounts are mostly qualitative, and the strongest causal estimates come from cross-national data. On balance, disinformation looks like a threat multiplier acting on older ethno-religious, political and economic grievances. The review also finds that the state shapes the threat it claims to fight. Slow or evasive communication by security agencies leaves vacuums that rumour fills, and laws built on vague offences, together with the 2021 Twitter suspension, have weakened the trust on which official correction depends. The 2022 NITDA Code of Practice points towards platform co-regulation, though its enforcement is untested. A research agenda closes the paper.

Keywords: Disinformation, National Security, Crisis Communication, Security Governance, Social Media, Nigeria

© 2026 The Authors. Published by Bexford Publishing Ltd

1. INTRODUCTION

Few Nigerians now get their political news from an editor. They get it from feeds. That shift from gatekept broadcasting to algorithmic curation changed who decides what counts as news (Casero-Ripollés, 2018; Papathanassopoulos & Giannouli, 2025). Platforms are not neutral pipes. Ranking systems reward engagement and advertising tools allow narrow targeting, so commercial logic ends up shaping political speech in ways users seldom notice (Papathanassopoulos & Giannouli, 2025). Much of this news arrives incidentally, in the middle of scrolling for something else, and incidental exposure tends to come with less scrutiny (Casero-Ripollés, 2018).

Deliberate falsehood has also become cheap to produce. On Twitter, false stories reached people roughly six times faster than true ones in Vosoughi et al.'s (2018) analysis, a figure that later reviews still lean on (Surjatmodjo et al., 2024). Generative AI lowers the cost again. A small operation can now make convincing synthetic video or audio, or recycle real footage under a false caption (Huntsman, 2025). Bot and troll networks then do the work of making a fringe claim look popular (Caled & Silva, 2022; Plikynas et al., 2025).

Is any of this a security problem, or simply a problem of poor media? A growing literature argues the former and treats information manipulation as a strategic activity aimed at belief and identity (Sługocki & Sowa, 2021). Surjatmodjo et al. (2024) reviewed 150 studies: 73% reported declining trust in government institutions, 67% reported attempts to manipulate opinion during elections, and 52% linked disinformation to intergroup conflict. Those percentages need careful reading. They count what researchers chose to study. They do not measure harm, and they say little about when an online falsehood becomes an offline one.

Nigeria offers plenty of material for that question. Ethnic and religious rivalries run deep, the economy is under strain, and institutional capacity is uneven (Hassan, 2023). Mobile internet has spread faster than digital literacy. Hassan's (2023) interviews and focus groups in all six geopolitical zones also show that falsehood rarely stays online. Rumours move onto partisan radio and television, often through paid political talkers known in Hausa as *Sojojin Baci* ("soldiers of mouth"), and from there into everyday conversation. Elections have been frequent targets. In 2019, doctored images circulated claiming that President Buhari had died and been replaced by a Sudanese double called "Jubrin" (Hassan, 2023). In the 2023 Lagos governorship race the slogan "Lagos is Yoruba land" was turned against Igbo residents, and Ademosu et al. (2026) link it to the intimidation of voters.

International research is poorly placed to explain cases like these. Much of it studies Western elections, above all the 2016 US contest, or builds machine-learning detectors from platform data (Plikynas et al., 2025). Little of it follows falsehood from phones to radio to market conversation in sub-Saharan Africa, and almost none asks how weak crisis communication and heavy-handed law interact with that flow. This review takes up that gap. Rather than assume that disinformation is a security threat by default, it asks where the evidence links online falsehood to security harm and where the claim outruns the data. The difference has practical weight: policy built on an inflated threat tends to produce laws that curb lawful speech without making anyone safer. Sections 2 to 5 cover concepts, the Nigerian threat literature, crisis communication and governance; Sections 6 to 8 offer a synthesis, a research agenda and conclusions.

2. CONCEPTUAL AND ANALYTICAL FOUNDATIONS

Disinformation, national security, crisis communication and security governance are usually studied in separate literatures. Figure 1 arranges them as a loop. I treat each link in it as a claim to be tested in Sections 3 to 5, not as an established fact.

© 2026 The Authors. Published by Bexford Publishing Ltd

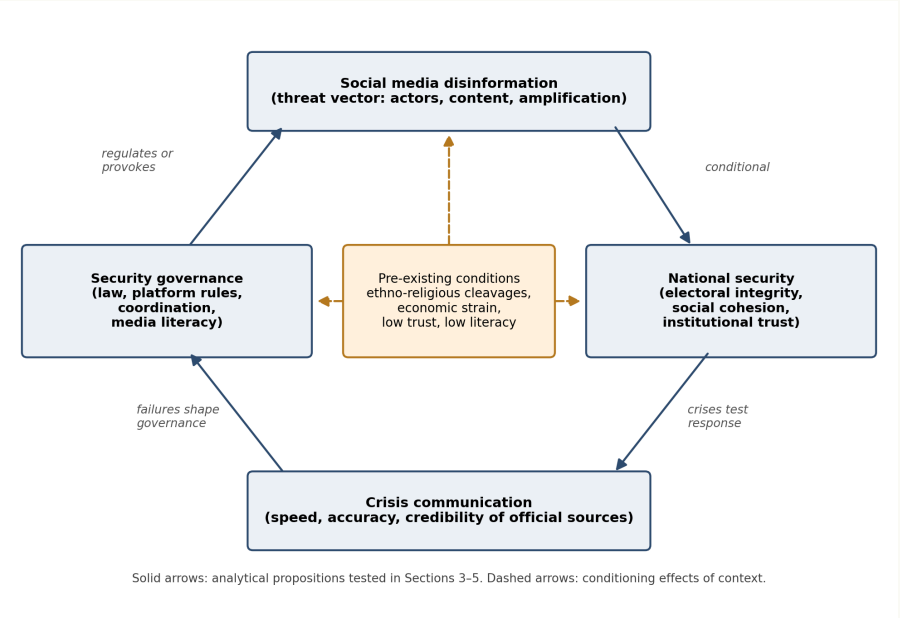


Figure 1. Conceptual framework linking disinformation, national security, crisis communication and security governance.

2.1 Social Media Disinformation

Wardle and Derakhshan's (2017) typology is still the usual starting point. Misinformation is falsehood shared without intent to harm; disinformation is falsehood made or spread on purpose; malinformation is true material used out of context to cause harm. For security analysis the middle category matters most, because intent points to actors and to accountability. Hameleers (2023) sharpens the definition by describing disinformation as covert, context-bound deception that can work by de-contextualising or manipulating real material as easily as by inventing it. Computational studies go another way. They fold fake news, propaganda and disinformation into one category and measure it with classifiers, bot scores and network metrics (Plikynas et al., 2025). The two traditions do not sit comfortably together. A classifier can flag falsity or coordination; it cannot read intent.

Mechanisms of spread are better documented. Fake accounts manufacture apparent support (Caled & Silva, 2022), and recommendation systems combined with homophily sort users into like-minded clusters (Plikynas et al., 2025). Demand matters as much as supply. Repetition makes claims feel true, and people favour claims that fit what they already believe (Caled & Silva, 2022). The once-popular idea that corrections backfire has not held up well in large replications (Wood & Porter, 2019). Alves Ferreira et al. (2026) go further and argue that distrust and existential uncertainty make simplified, conspiratorial stories attractive. If susceptibility is partly social, purely technical fixes will miss part of it.

2.2 National Security in the Digital Information Environment

Security studies now extend to the information environment, and some authors speak of cognitive security, meaning protection of the processes through which citizens form beliefs and judge authority (Sługocki & Sowa, 2021). The vulnerabilities that recur in this work are electoral legitimacy, cohesion between groups and trust in institutions (Surjatmodjo et al., 2024). Nigeria adds an aggravating feature. A presidential system that personalises power, combined with politics organised around identity, gives character attacks and ethnic narratives an unusually receptive audience (Ademosu et al., 2026; Hassan, 2023).

The wider concept carries a cost. Once information becomes a security domain, governments gain a vocabulary for treating satire, dissent or plain error as threats. Using the concept responsibly means showing that a particular campaign actually damages trust, cohesion or electoral legitimacy.

© 2026 The Authors. Published by Bexford Publishing Ltd

2.3 Crisis Communication and Disinformation

Crises compress time. When official messages are slow, vague or defensive, rumour fills the gap, and some of that rumour is planted (Sługocki & Sowa, 2021). Sługocki and Sowa treat pandemic disinformation as a test of national security capacity, and the wider literature agrees that late, dismissive rebuttals tend to fail; prompt, transparent messaging and prebunking do better (Caled & Silva, 2022). Nigeria complicates a model that assumes a digital battlefield. Rumours there pass from WhatsApp to radio and then to face-to-face networks, while the police keep only a thin online presence (Hassan, 2023). An agency missing from both arenas has little say in how a crisis is understood.

2.4 Security Governance in the Digital Information Environment

Governance responses split along a familiar line. One model gives the executive broad powers to label and punish falsehood; Nigeria's Protection from Internet Falsehood and Manipulation Bill (2019) and Singapore's Protection from Online Falsehoods and Manipulation Act are examples. Critics warn that such powers threaten the expression protected by Section 39 of Nigeria's Constitution and invite use against opponents (Chiamogu et al., 2025; Hassan, 2023). The other model spreads responsibility across agencies, fact-checkers, civil society, universities and platforms (Caled & Silva, 2022).

Evidence on particular tools is uneven. Media literacy has the most consistent support. Fact-checking and inoculation help less reliably (Caled & Silva, 2022), and automated detection errs in both directions, particularly outside English (Huntsman, 2025). Decentralised protocols such as ActivityPub and the AT Protocol move moderation away from single companies, though whether that reduces disinformation is not yet known (Ciriello et al., 2026; Oshinowo et al., 2026). Governance thus closes the loop by acting back on the threat, in no fixed direction. Literacy and credible institutions may shrink demand for falsehood. Criminalising speech may deepen the distrust that falsehood feeds on.

3. SOCIAL MEDIA DISINFORMATION AS A NATIONAL SECURITY THREAT IN NIGERIA

3.1 Evolution and Forms of Social Media Disinformation

Nigerian disinformation is older than social media, but its production has changed. Olanipekun (2025) describes a move from broadcast propaganda aimed at mass audiences to computational propaganda that targets individuals by their online behaviour. Synthetic media, and bot networks that push hashtags into trending lists, dominate this literature (Olanipekun, 2025; Surjatmodjo et al., 2024). The Nigerian studies care less about technical sophistication. What Hassan (2023) documents is movement between media: broadcasters under commercial pressure picking up unverified posts, and *Sojojin Baci* mixing fact and invention on air. The most damaging falsehood in Nigeria may well be a low-tech rumour that borrows a trusted presenter's voice.

Content has turned towards identity. Ademosu et al. (2026) found 2023 Lagos campaign messaging that combined ordinary acclaim and attack with ethnic triggers, setting "Lagos is Yoruba land" against "Lagos is no man's land." The actors range from governments and party "cyber armies" to armed groups such as Boko Haram and IPOB, commercial consultants and ordinary partisans (Hassan, 2023; Piazza, 2022). That range matters, since a foreign campaign and a forwarded rumour call for different responses.

3.2 Pathways Through Which Disinformation Creates Security Risks

Proposed pathways from content to harm operate at different levels. For individuals, disinformation trades on alarm (emergency, betrayal, catastrophe), and repetition breeds a familiarity that readers mistake for credibility (Alves Ferreira et al., 2026; Olanipekun, 2025). Economic strain may add to this by making conspiratorial explanations comforting (Alves Ferreira et al., 2026).

© 2026 The Authors. Published by Bexford Publishing Ltd

Piazza (2022) works at the level of the political system. Using data on more than 150 countries between 2000 and 2017, he argues that disinformation spread by political actors polarises societies and that polarisation in turn raises the risk of domestic terrorism. His mediation estimates repay close reading. Polarisation carries between 7.1% and 24.3% of the effect, depending on the actor, which makes it a real pathway but a partial one; most of the effect travels through channels the model does not name. One candidate he offers is the collapse of pluralistic ignorance, as platforms show people with extreme views that they are not alone.

Nigeria adds an institutional pathway. Where security agencies are absent online, armed groups speak to the public directly. Hassan (2023) records Boko Haram and IPOB circulating atrocity footage, including executions of captured soldiers, to intimidate and to provoke reprisal. The danger lies partly in the content and partly in the missing competitor.

3.3 Disinformation, Social Instability and Security Crises

Evidence on instability is broad but uneven. The steadiest findings concern trust. Repeated exposure to false narratives erodes confidence in news, officials and science, with knock-on effects for compliance in emergencies (Alves Ferreira et al., 2026; Surjatmodjo et al., 2024). Identity-targeted content deepens polarisation (Ademosu et al., 2026; Piazza, 2022), and sustained campaigns coincide with lower faith in bodies such as INEC (Hassan, 2023).

Violence is documented episodically. Fabricated videos claiming that food had been poisoned with pesticides inflamed relations between Hausa and Igbo communities, and false reports of attacks have prompted reprisals in neighbouring areas (Hassan, 2023). On election days, rumours of shootings have reportedly kept voters at home (Ademosu et al., 2026; Hassan, 2023), though nobody has measured the size of that effect. Escalation from online claim to street disorder is not unique to Nigeria; the Jakarta post-election riots and the attack on the US Capitol followed a similar path (Piazza, 2022; Surjatmodjo et al., 2024). Terrorism is the one outcome with cross-national causal evidence (Piazza, 2022), and Hassan (2023) shows insurgents in Nigeria using platforms for propaganda and recruitment. Neither finding shows that disinformation drives Nigeria's insurgencies.

3.4 Nigerian Vulnerabilities and Contextual Factors

Nigeria's exposure comes from conditions acting together. More than 250 ethnic groups and long-mismanaged regional and religious rivalries supply the cleavages (Ademosu et al., 2026; Hassan, 2023). Elections are high-stakes, and politicians hire social media teams to circulate false claims and ethnic slurs (Hassan, 2023). Two features stand out. One is the broadcast pipeline that carries online falsehood to rural listeners in Hausa, Yoruba and English. The other is the near-silence of the police and the Nigeria Security and Civil Defence Corps on digital channels, which leaves people relying on peer-to-peer rumour for security news (Hassan, 2023). The state's legislative answer, the 2019 Internet Falsehood Bill, proposed sweeping powers and did nothing about why falsehood spreads (Hassan, 2023).

Neighbouring countries show similar patterns, from WhatsApp rumour networks in close Sierra Leonean elections to party "cyber battalions" in Ghana and ethnic propaganda tied to post-election violence in Kenya (Hassan, 2023). What may set Nigeria apart is how many of these conditions coexist.

3.5 Critical Synthesis of the Threat Literature

Sorting this literature by what each study can actually show gives a more modest picture than the security framing suggests (Table 1). Causal evidence is scarce. Piazza's (2022) instrumental-variable estimates depend on the validity of his instruments and describe averages across many countries, not mechanisms inside Nigeria. Amplification is better established (Surjatmodjo et al., 2024; Vosoughi et al., 2018). Much

© 2026 The Authors. Published by Bexford Publishing Ltd

of the rest is correlational or theoretical; the existential-disconnection account offered by Alves Ferreira et al. (2026) is a proposal rather than a finding.

Table 1. The threat literature classified by type and strength of evidence

Type of evidence	Claim	Principal sources	Assessment
Causal	Disinformation by political actors raises domestic terrorism	Piazza (2022)	Strong design; cross-national, silent on Nigeria
Amplification	False content spreads faster and is enlarged by bots, algorithms and broadcasters	Vosoughi et al. (2018); Surjatmodjo et al. (2024); Hassan (2023)	Strong and consistent
Correlational	Exposure goes with lower trust, polarisation and lower turnout	Surjatmodjo et al. (2024); Ademosu et al. (2026)	Moderate; direction unclear
Qualitative (Nigeria)	Rumours cross into broadcasting; elections and communal tensions are targeted	Hassan (2023); Ademosu et al. (2026)	Rich but perception-based
Theoretical	Existential uncertainty and boredom raise susceptibility	Alves Ferreira et al. (2026)	Untested in Nigeria
Speculative	Disinformation is the main driver of insurgency	None of the reviewed studies	Unsupported

Source: Author's classification of the reviewed literature.

The Nigerian evidence is mostly qualitative. Hassan's (2023) interviews record perceptions rather than measured effects, and Ademosu et al. (2026) analysed a small purposive sample of posts. Platform opacity blocks independent audits of the algorithms on which the amplification argument rests (Surjatmodjo et al., 2024). A quieter disagreement concerns who the main actors are. Western work has centred on foreign interference, whereas Nigerian studies put domestic elites, cyber armies and broadcasters at the heart of the problem (Ademosu et al., 2026; Hassan, 2023).

On this evidence, disinformation is best described as a threat multiplier in Nigeria. It accelerates conflicts rooted in poverty, corruption, governance failure and impunity within the security forces. Claims that it drives insurgency or state fragility on its own go beyond what the studies show (Hassan, 2023; Piazza, 2022).

4. DISINFORMATION AND CRISIS COMMUNICATION

4.1 Information Disorder During Security Crises

Security crises raise anxiety, weaken scrutiny and create demand for information faster than officials can meet it (Shahbazi & Bunker, 2024). Smartphones have also ended the state's monopoly on first accounts. In Borno, Kaduna, Zamfara and Plateau, witness footage now circulates before any agency speaks. It

© 2026 The Authors. Published by Bexford Publishing Ltd

builds informal early-warning networks; it also spreads recycled and doctored images (Ariwodola & Tijani, 2026; Ya'u & Usman, 2026). The two effects are hard to pull apart. Hilberts et al.'s (2025) narrative review of nine disaster studies found that misinformation disrupted emergency response, but that evidence comes from floods and hurricanes. Comparable studies of Nigerian security incidents are scarce.

4.2 Government and Institutional Communication

Nigerian research on security communication leans on Situational Crisis Communication Theory, which ties the right response to how much responsibility the public assigns (Coombs, 2007), and on Benoit's (1997) image repair strategies. Agencies have tended towards the defensive end. Studies report late releases, little transparency and conflicting accounts from Defence Headquarters, the police, the Department of State Services and the Ministry of Information (Ariwodola & Tijani, 2026; Garba, 2024). Some delay reflects a real dilemma, since early statements risk error. Sievi and Pawelec (2025) suggest prompt updates that say plainly what is and is not confirmed, an idea not yet tested in Nigeria.

The 2021 attack on the Nigerian Defence Academy shows what silence costs. Two officers died and a major was abducted. Defence Headquarters posted about unrelated meetings and issued one statement attacking a news outlet's CCTV report. Uwalaka (2026) traces the result through more than 8,000 Facebook posts and comments, many demanding the removal of the service chiefs. Ya'u and Usman (2026) describe the contest beneath such episodes. Police statements frame isolated crimes under control; social media users frame state failure.

4.3 Public Trust and Institutional Credibility

Credibility is what these contests are about. Dei (2025) argues that campaigns against courts and police aim at cognitive insecurity, a state in which citizens stop believing that institutions are neutral, and he identifies frames that present institutions as partisan tools, as corrupt, or as threats to particular groups. His analysis is conceptual and not Nigerian, yet each frame is familiar here. Shahbazi and Bunker (2024) supply a mechanism: contradictory official messages break shared situational awareness and send people to unofficial sources. Garba (2024) finds Nigerian security communication putting reputation ahead of disclosure, and Balaleye and Adelakun (2024) report lasting damage to police credibility after #EndSARS. Official behaviour can therefore feed the problem it is meant to solve.

4.4 Counter-Disinformation Communication

Most countermeasure research comes from elsewhere. Prebunking, which warns audiences about manipulation tactics in advance, improved discernment in a large field study on YouTube (Roozenbeek et al., 2022), and warnings about misinformation look promising in vaccine research (Whitehead et al., 2023). In US and Dutch experiments, literacy messages alone lowered belief in false claims without shifting agreement with them; pairing a forewarning with a fact-check moved both (Hameleers, 2022). Wagnsson et al. (2025) found that literacy training framed around foreign threats avoided breeding cynicism about reliable media, but their setting was a high-trust society that Nigeria does not resemble. The messenger matters as well. Sievi and Pawelec (2025) caution that security agencies running public education may look paternalistic or partisan, a serious risk where police credibility is already low. Whitehead et al. (2023) also found that scare tactics and overconfident messages backfire. Whether these lessons hold in Nigerian security crises remains open.

5. SECURITY GOVERNANCE AND NIGERIA'S RESPONSE

5.1 Government and Security Institutions

The state's approach has been driven less by disinformation than by a wider securitisation of political

© 2026 The Authors. Published by Bexford Publishing Ltd

expression. Drawing on securitisation theory (Buzan et al., 1998), Chiamogu et al. (2025) and Omede and Oladimeji (2025) show executive elites presenting #EndSARS and #RevolutionNow as existential threats, a move that justified arrests and platform suspension (Kojah et al., 2025). Assertiveness about speech coexists with a weak voice. Security agencies remain reactive online, leaving room for Boko Haram, ISWAP and IPOB, and the National Orientation Agency and the information ministry have little digital reach and are widely seen as partisan (Ezeorah et al., 2025; Hassan, 2023). Put bluntly, the state polices speech more than it informs. Omede and Oladimeji's (2025) Lagos survey hints at the consequence: online documentation of abuses went with greater accountability and also with a wider trust gap between citizens and security agencies. It is one city and one episode, but the pattern deserves attention.

5.2 Legal and Regulatory Responses

Nigerian law has been primarily concerned with users, viewing citizens and journalists as publishers accountable for their posts (Inobemhe & Ossai, 2026; Obia, 2025). Section 24 of the Cybercrimes (Prohibition, Prevention, etc.) Act 2015 criminalized grossly offensive or false messages (Chiamogu et al., 2025). On 25 March 2022, the ECOWAS Court of Justice ruled that the section was incompatible with freedom of expression and ordered its amendment (Socio-Economic Rights and Accountability Project v. Federal Republic of Nigeria, 2022) . Parliament rewrote it in 2024 (Cybercrimes (Prohibition, Prevention, etc.) (Amendment) Act, 2024), and whether practice has changed has not been studied. Bills on internet falsehood and hate speech stalled amid public opposition (Inobemhe & Ossai, 2026).

There is a broad consensus among commentators that these provisions are overly broad and have been applied against journalists and critics (Inobemhe & Ossai, 2026; Obia, 2025). The asymmetry is telling, according to Chiamogu et al. (2025): the executive backed restrictive bills, but President Buhari refused to sign the Digital Rights and Freedom Bill. By contrast, Twitter's seven-month ban in June 2021 on deleting a presidential post did not even go through the courts (Chiamogu et al., 2025; Kojah et al., 2025). Nigeria is hardly unique in this. Uganda shut down the internet during the 2021 election and Ethiopia issued a wide-ranging proclamation on hate speech in 2020 (Kojah et al., 2025; Obia, 2025). Figure 2 sets these measures in sequence. Very little of this law addresses the coordinated, identity-based manipulation described in Section 3.

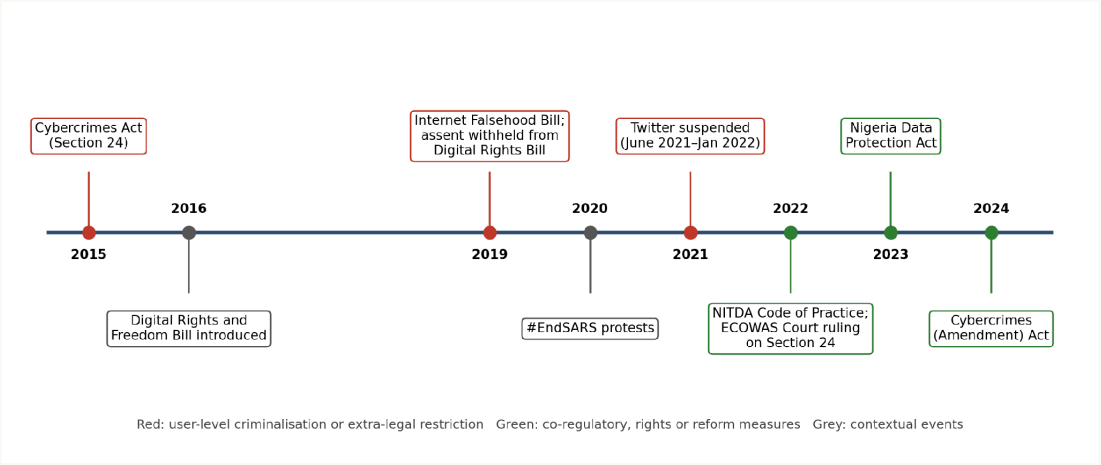


Figure 2. Major legal, regulatory and extra-legal measures affecting Nigeria's digital information environment, 2015–2024.

Source: Author's construction from the sources and legislation cited.

5.3 Platform Governance

The NITDA Code of Practice for Interactive Computer Service Platforms, issued in 2022, changes

© 2026 The Authors. Published by Bexford Publishing Ltd

direction. Obia (2025) calls it a co-regulatory turn because it places duties of care on platforms rather than users. Platforms must remove unlawful content within 48 hours of a substantiated notice, keep records and file annual compliance reports, and Part V asks them to support local fact-checkers and give researchers access to data (Obia, 2025). On paper it resembles the EU's Digital Services Act. Nigeria, however, lacks the EU's market leverage, and Obia (2025) finds no documented enforcement or compliance reporting by the major platforms.

Self-governance does not fill the gap. Warnke et al. (2024) show that Facebook, YouTube and Twitter kept their COVID-19 responses within limits set by their own operating principles. Moderation in Hausa, Yoruba and Igbo is thin (Hassan, 2023). Kojah et al. (2025) document the reverse failure, with automated systems suppressing posts by marginalised users and human rights defenders. Harmful local content is under-moderated while accountability content is over-moderated.

5.4 Institutional Coordination

Four regulators share the field: NITDA, the Nigerian Communications Commission, the National Broadcasting Commission and the Nigerian Press Council (Inobemhe & Ossai, 2026; Obia, 2025). Their mandates overlap and their laws conflict. Part VI of the NITDA Code borrows penalties from statutes that allow imprisonment of corporate officers, although Section 3 of the Code rules out individual criminal liability, and its data-retention rules clash with erasure rights under the Nigeria Data Protection Act 2023 (Obia, 2025). On multi-stakeholder coordination, the clearest Nigerian evidence comes from health. Ezenwaka et al.'s (2025) review of the COVID-19 response found agencies working in silos without shared data. Nothing in the reviewed literature suggests that information governance does better.

5.5 Governance Gaps and Implementation Challenges

These gaps compound one another. Legally, undefined terms such as "grossly offensive" invite arbitrary enforcement, and the Code rests on executive authority without legislative backing (Inobemhe & Ossai, 2026; Obia, 2025). NITDA depends on platforms reporting on themselves, and security agencies lack staff trained in digital communication and open-source verification (Ezeorah et al., 2025; Obia, 2025). Legitimacy may be the deepest problem. Because earlier measures were turned on critics, many citizens read any new anti-disinformation initiative as an attempt to silence dissent (Chiamogu et al., 2025). The costs are visible too. Prosecutions and bans encourage self-censorship and push conversation into encrypted channels, where falsehood is harder to see or answer (Chiamogu et al., 2025; Hassan, 2023). Those costs are better documented than any reduction in disinformation, which nobody has yet measured (Obia, 2025).

6. CRITICAL SYNTHESIS: FROM INFORMATION DISORDER TO SECURITY GOVERNANCE

Taken together, the evidence supports the chain set out in the introduction only in part. Its early links hold: platforms reward emotive content, and false claims outrun true ones (Surjatmodjo et al., 2024; Vosoughi et al., 2018). The middle links are conditional. Exposure goes with lower trust and sharper polarisation, but mostly where distrust, hardship and identity cleavage already exist (Alves Ferreira et al., 2026; Hassan, 2023). The final link, to violence, rests on Nigerian anecdote and cross-national estimates (Piazza, 2022).

Two features set Nigeria apart from the Western literature. Rumours gain authority when broadcasters repeat them and travel further through closed WhatsApp groups, so platform-centred models miss much of the problem (Cheeseman et al., 2020; Hassan, 2023). The main actors are also domestic. Interventions built around foreign threats, such as the threat-primed literacy training tested by Wagnsson et al. (2025), are therefore an uncertain fit.

The most important finding concerns the state. The introduction treated crisis communication and

© 2026 The Authors. Published by Bexford Publishing Ltd

governance as responses to an external threat, but the evidence suggests they are also part of the causal process (Figure 3). Silence leaves vacuums for rumour and armed groups (Hassan, 2023; Uwalaka, 2026). Contradictory statements push citizens towards unofficial sources (Shahbazi & Bunker, 2024). Prosecutions and bans confirm suspicions that anti-disinformation policy is about control (Chiamogu et al., 2025; Obia, 2025). Each erodes the trust that official correction needs.

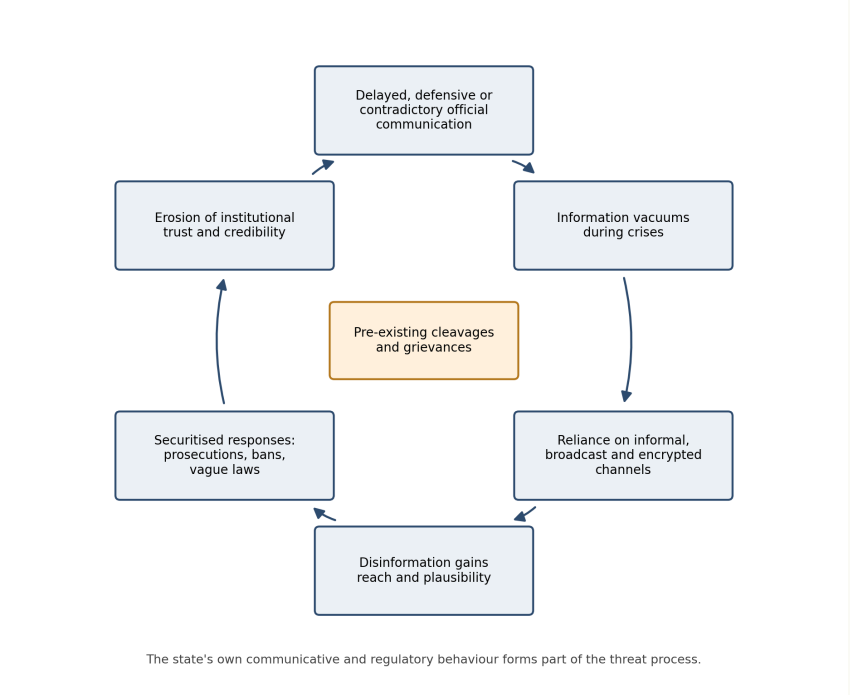


Figure 3. The credibility trap: how state communication and regulatory responses can reinforce the disinformation threat.

Source: Author's synthesis of Hassan (2023), Shahbazi and Bunker (2024), Chiamogu et al. (2025), Obia (2025) and Uwalaka (2026).

Nigeria's difficulty is less a shortage of legal power than a shortage of credibility. The state has ample tools to punish speech and is largely absent from the contest to inform. More coercive capacity without rebuilt trust would tighten the loop in Figure 3. Transparent crisis communication, local-language moderation, combined prebunking and fact-checking, and enforceable co-regulation all have support in the literature (Hameleers, 2022; Obia, 2025; Roozenbeek et al., 2022), though almost none of that support comes from Nigeria.

The literacy case illustrates the problem. Surjatmodjo et al. (2024) rate digital literacy as the most effective countermeasure in the 150 studies they reviewed, yet 87% of those studies were international or global in scope, and few deal with low-trust, multilingual settings. A programme that works in a classroom in Europe may fail on a WhatsApp group in Kano, especially if it arrives with a government logo attached. Before such programmes are scaled up in Nigeria, someone needs to test who should deliver them and in which language.

7. RESEARCH GAPS AND FUTURE RESEARCH AGENDA

Five gaps stand out. Closed messaging is the first. WhatsApp is central to Nigerian politics (Cheeseman et al., 2020; Hassan & Hitchen, 2022), yet most evidence comes from public platforms and interviews; tiplines and voluntary data donation, with privacy safeguards, could open closed groups to study. Second, nobody has measured how much online falsehood reaches radio. Monitoring Hausa, Yoruba, Igbo and Pidgin broadcasts alongside platform data would let specific claims be traced. Third, mechanisms of harm

© 2026 The Authors. Published by Bexford Publishing Ltd

need subnational designs. Matching geolocated conflict events from the Armed Conflict Location and Event Data Project (Raleigh et al., 2010) to the timing of rumours could show when online claims precede violence and when they do not. Fourth, interventions need testing in Nigerian languages, including whether security agencies, civil society or community figures make more credible messengers. Fifth, governance outcomes remain unexamined: platform compliance with the NITDA Code, prosecutions under the amended Section 24, and the accuracy of automated moderation in Nigerian languages (Kojah et al., 2025; Obia, 2025). Transparency reports, court records and audit studies could answer each of these.

8. CONCLUSION

The question this review set out to answer was when, rather than whether, social media disinformation becomes a security concern in Nigeria. The answer is qualified. False content spreads quickly and gains force by moving from phones into broadcasting and conversation. Where it meets ethnic and religious fault lines, contested elections or acute insecurity, it can sharpen fear and hostility. The literature does not show it to be the main driver of Nigeria's insecurity; it acts on grievances that exist without it.

The state is part of the story. Silence in crises, contradictory messaging, vague criminal provisions and the Twitter suspension have each weakened the credibility of official information, which is the main resource a government has for correcting falsehood. The NITDA Code shows that another approach is possible, but its enforcement is untested.

Policy built on this evidence would put credibility before control. That means candid and timely communication by security agencies, coordination across institutions, moderation capacity in Nigerian languages, legislative grounding for co-regulation, and literacy work delivered by trusted intermediaries. None of these has been rigorously evaluated in Nigeria. Until they are, claims about the size of the threat and the value of particular remedies deserve caution.

References

Ademosu, I., Aondover, E. M., Osadolor, B., & Abraham, B. (2026). Divisive politics: Analysis of ethnic rhetoric and hate speech in the 2023 Lagos Governorship elections. *Frontiers in Political Science, 7*, Article 1663797. https://doi.org/10.3389/fpos.2025.1663797

Alves Ferreira, J. M., Giannouli, V., & Tukaiev, S. (2026). Disinformation, psychosocial vulnerability, and media trust in the digital era: Implications for health behaviour and societal resilience. *Healthcare, 14*(8), Article 1089. https://doi.org/10.3390/healthcare14081089

Ariwodola, M. A., & Tijani, A. A. (2026). Appraisal of crisis communication on national security in Nigeria: A conceptual gap analysis. *Journal of Arts and Sociological Research, 11*(6), 24–36. https://doi.org/10.70382/ajasr.v11i6.097

Balaleye, T., & Adelakun, J. G. (2024). Assessment of public relations strategies in crisis management during the 2020 #EndSARS protest and Lekki shooting in Lagos, Nigeria. *Journal of Communication and Media Technology, 4*(1).

Benoit, W. L. (1997). Image repair discourse and crisis communication. *Public Relations Review, 23*(2), 177–186. https://doi.org/10.1016/S0363-8111(97)90023-0

Buzan, B., Wæver, O., & de Wilde, J. (1998). *Security: A new framework for analysis*. Lynne Rienner.

Caled, D., & Silva, M. J. (2022). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science, 5*(1), 123–159.

© 2026 The Authors. Published by Bexford Publishing Ltd

https://doi.org/10.1007/s42001-021-00118-8

Casero-Ripollés, A. (2018). Research on political information and social media: Key points and challenges for the future. *El Profesional de la Información, 27*(5), 964–974. https://doi.org/10.3145/epi.2018.sep.01

Cheeseman, N., Fisher, J., Hassan, I., & Hitchen, J. (2020). Social media disruption: Nigeria's WhatsApp politics. *Journal of Democracy, 31*(3), 145–159. https://doi.org/10.1353/jod.2020.0037

Chiamogu, A. P., Obikeze, O. S., Chiamogu, U. M., & Odikpo, C. (2025). Social media regulation and autocratisation in Nigeria: Dialectics of shrinking civic space under Buhari's administration. *African Journal of Politics and Administrative Studies, 13*(1), 199–234.

Ciriello, R. F., Marx, J., Cheong, M., Mueller-Bloch, C., & Mathiassen, L. (2026). Decentralized social media. *Business & Information Systems Engineering, 68*(1), 219–234. https://doi.org/10.1007/s12599-025-00952-4

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review, 10*(3), 163–176. https://doi.org/10.1057/palgrave.crr.1550049

Cybercrimes (Prohibition, Prevention, etc.) (Amendment) Act, 2024 (Nigeria).

Dei, M. (2025). Cognitive security and institutional trust: A sociological analysis of disinformation campaigns targeting the legitimacy of judicial and law enforcement institutions. *Politics & Security, 14*(4), 45–56. https://doi.org/10.54658/ps.28153324.2025.14.4.pp.45-56

Ezenwaka, U., Mbachu, C., & Onwujekwe, O. (2025). A scoping review of the roles of stakeholders and coordination mechanisms for enhanced multi-sectoral and multi-level interventions in COVID-19 response in Nigeria. *Health Research Policy and Systems, 23*, Article 18. https://doi.org/10.1186/s12961-024-01276-7

Ezeorah, C. E., Luka, R. C., & Erunke, C. (2025). Effects of social media on national security in Nigeria, 2015–2024. *The International Journal of Social Sciences Research and Anthropology, 7*(6), 286–304. https://doi.org/10.70382/tijssra.v07i6.035

Garba, S. (2024). Public relations in postcrisis recovery: Analysing Nigeria's government communication in the security sector. *African Journal of Social and Behavioural Sciences, 14*(5).

Hameleers, M. (2022). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society, 25*(1), 110–126. https://doi.org/10.1080/1369118X.2020.1764603

Hameleers, M. (2023). Disinformation as a context-bound phenomenon: Toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory, 33*(1), 1–10. https://doi.org/10.1093/ct/qtac021

Hassan, I. (2023). Dissemination of disinformation on political and electoral processes in Nigeria: An exploratory study. *Cogent Arts & Humanities, 10*(1), Article 2216983. https://doi.org/10.1080/23311983.2023.2216983

Hassan, I., & Hitchen, J. (Eds.). (2022). *WhatsApp and everyday life in West Africa: Beyond fake news*. Bloomsbury Academic.

© 2026 The Authors. Published by Bexford Publishing Ltd

Hilberts, S., Govers, M., Petelos, E., & Evers, S. (2025). The impact of misinformation on social media in the context of natural disasters: Narrative review. *JMIR Infodemiology, 5*, Article e70413. https://doi.org/10.2196/70413

Huntsman, L. H. (2025). Editorial: Disinformation countermeasures and artificial intelligence. *Frontiers in Artificial Intelligence, 8*, Article 1750972. https://doi.org/10.3389/frai.2025.1750972

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. https://doi.org/10.3126/ajhss.v3i1.92792

Kojah, S. A., Zhang, B. Z., Are, C., Delmonaco, D., & Haimson, O. L. (2025). Silencing the voiceless: Social media content moderation for at-risk marginalized populations in high-stakes violence and human rights contexts in Nigeria. *Proceedings of the ACM on Human-Computer Interaction, 9*(CSCW2), Article CSCW323. https://doi.org/10.1145/3757504

Obia, V. (2025). Digital policy and Nigeria's Platform Code of Practice: Towards a radical co-regulatory turn. *Data & Policy, 7*, Article e12. https://doi.org/10.1017/dap.2025.1

Olanipekun, S. O. (2025). Computational propaganda and misinformation: AI technologies as tools of media manipulation. *World Journal of Advanced Research and Reviews, 25*(1), 911–923. https://doi.org/10.30574/wjarr.2025.25.1.0131

Omede, A. J., & Oladimeji, O. B. (2025). Effect of social media on national security during the 2020 ENDSARS protests in Lagos State. *De-centre: Journal of Interdisciplinary Studies, 1*(1), 1–13.

Oshinowo, T., Hwang, S., Zhang, A. X., & Monroy-Hernández, A. (2026). Seeing the politics of decentralized social media protocols. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '26)*. Association for Computing Machinery.

Papathanassopoulos, S., & Giannouli, I. (2025). Political communication in the age of platforms. *Encyclopedia, 5*(2), Article 77. https://doi.org/10.3390/encyclopedia5020077

Piazza, J. A. (2022). Fake news: The effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict, 15*(1), 55–77.

Plikynas, D., Rizgelienė, I., & Korvel, G. (2025). Systematic review of fake news, propaganda, and disinformation: Examining authors, content, and social impact through machine learning. *IEEE Access, 13*. https://doi.org/10.1109/ACCESS.2025.3530688

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research, 47*(5), 651–660. https://doi.org/10.1177/0022343310378914

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances, 8*(34), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Shahbazi, M., & Bunker, D. (2024). Utilising social media for shared situational awareness (SSA) during crises: Link quality, communication strategies and trust. *International Journal of Information Management, 77*, Article 102780. https://doi.org/10.1016/j.ijinfomgt.2024.102780

Sievi, L., & Pawelec, M. (2025). (How) should security authorities counter false information on social media in crises? A democracy-theoretical and ethical reflection. *International Journal of Disaster*

© 2026 The Authors. Published by Bexford Publishing Ltd

Risk Reduction, 116, Article 105093. https://doi.org/10.1016/j.ijdrr.2024.105093

Socio-Economic Rights and Accountability Project v. Federal Republic of Nigeria, Application No. ECW/CCJ/APP/09/19 (ECOWAS Community Court of Justice, March 25, 2022).

Surjatmodjo, D., Unde, A. A., Cangara, H., & Sonni, A. F. (2024). Information pandemic: A critical review of disinformation spread on social media and its implications for state resilience. *Social Sciences, 13*(8), Article 418. https://doi.org/10.3390/socsci13080418

Sługocki, W. Ł., & Sowa, B. (2021). Disinformation as a threat to national security on the example of COVID-19 pandemic management. *Security & Defence Quarterly, 35*(3), 61–72. https://doi.org/10.35467/sdq/138876

Uwalaka, T. (2026). Nigerian military strategic use of social media during online firestorms: An appraisal of the NDA terrorist attack. *Journal of Communication Inquiry, 50*(3), 401–421. https://doi.org/10.1177/01968599231151727

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wagnsson, C., Östervall, A., & Angwald, A. (2025). Naming the enemy: How to fortify society against foreign disinformation while avoiding excessive vigilance to reliable media. *Humanities and Social Sciences Communications, 12*. https://doi.org/10.1057/s41599-025-04844-6

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Council of Europe Report DGI(2017)09). Council of Europe.

Warnke, L., Maier, A.-L., & Gilbert, D. U. (2024). Social media platforms' responses to COVID-19-related mis- and disinformation: The insufficiency of self-governance. *Journal of Management and Governance, 28*(4), 1079–1115. https://doi.org/10.1007/s10997-023-09694-5

Whitehead, H. S., French, C. E., Caldwell, D. M., Letley, L., & Mounier-Jack, S. (2023). A systematic review of communication interventions for countering vaccine misinformation. *Vaccine, 41*(5), 1018–1034. https://doi.org/10.1016/j.vaccine.2022.12.059

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior, 41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Ya'u, H. A., & Usman, F. U. (2026). Framing insecurity in North-West Nigeria: A conceptual analysis of social media actors and Nigeria Police Force narratives. *Azman University International Journal of Management and Social Sciences, 1*(1), 35–42. https://doi.org/10.67523/auijmss.v1i1.005

© 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/)