IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Social Sciences and African Studies

Vol 1, No 1, 2026

Social Media, Misinformation and Social Cohesion in Plural Societies: A Theoretical Account of Algorithmic Amplification, Ethnic Mobilisation and Trust in Africa

NWOGBOJI Emmanuel Usulor[1], ATOLAGBE Umar Kawu[2], IGWE Emmanuel Sunday[3]

[1,2,3] Veristzon Global Research Lab

ABSTRACT

Africa's plural societies manage ethnic, religious and regional difference through institutions, norms and everyday relationships that social media increasingly bypass, delivering content selected by engagement-based ranking and by networks of in-group trust. Research on misinformation, hate speech and political communication in Africa is growing but fragmented, and its focus on falsity obscures the harm done by divisive content that need not be false. This theoretical paper addresses that gap. Its objective is to explain how social media affect cohesion in plural societies, to specify when the outcome is polarisation and violence rather than noise, and to identify the factors and policies that moderate harm. Methodologically, the study is an integrative, theory-building review. The population was peer-reviewed journal articles published between 2023 and 2026; a purposive sample of 45 articles was retained after each was verified against its publisher record and screened for retractions and corrections. Findings were coded against provisional mechanisms, compared across countries and platforms, and synthesised abductively. The analysis yields an amplification–mobilisation–erosion model. Amplification describes what ranking systems and messaging networks select; mobilisation, how political entrepreneurs such as campaigns, paid influencers and diaspora networks exploit that selection; and erosion, what happens to inter-group and institutional trust. The paper argues that the principal harm is the systematic over-supply of identity-threat content to populations whose cohesion depends on its under-supply. Seven propositions specify the roles of engagement ranking, closed networks, high-stakes political moments and historical grievance, and of three resilience factors: cross-cutting ties and shared identity, trusted local intermediaries, and institutional credibility. Nigeria is the anchoring case. The paper recommends integrating information integrity into election security, co-regulation rather than platform bans, preference-based ranking and local-language moderation, support for trusted intermediaries and peer correction, and disclosure rules for paid political influence.

Keywords: social media; misinformation; disinformation; algorithmic amplification; ethnicity

1. INTRODUCTION

Africa's plural societies have long managed ethnic, religious and regional difference through institutions, norms and everyday relationships: federal bargains, power-sharing conventions, intermarriage, markets, congregations and the mediating authority of community and religious leaders. Social media now bypass much of this mediating architecture. They deliver to individual users, at scale and at speed, content selected by ranking systems and by networks of trust that reward attention rather than accuracy or restraint. In Nigeria, the anchoring case of this paper, hate speech on Twitter/X peaks during political periods and is dominated by political, ethnic, religious and regional themes (Ridwanullah et al., 2025); the 2023 presidential campaign saw candidates' media teams both produce and respond to misleading claims in a climate of ethnic and religious tension (Ibrahim & Ekdale, 2026); and voters report that exposure to inflammatory campaign rhetoric deepened divisions and eroded confidence in electoral processes (Aondover et al., 2026).

Public debate tends to frame the problem as "misinformation": false claims that could, in principle, be corrected. The research reviewed here suggests that this framing is too narrow. Much of the content that damages inter-group relations is not straightforwardly false. It is selective, emotive and identity-laden; it dramatises threat from an out-group; and it travels because it is engaging. An audit of engagement-based ranking found that it amplifies emotionally charged content hostile to the political out-group relative to a chronological feed, and that users do not in fact prefer the political content the algorithm selects (Milli et al., 2025). In closed messaging networks, emotional resonance and trust within groups drive the spread of disinformation more than factual accuracy, and corrections are dismissed as propaganda (Gondwe, 2025).

The literature on these phenomena in Africa is growing but fragmented. Discourse studies catalogue the rhetorical forms of hate speech; political-communication studies document campaigns, influencers and messaging groups; experimental studies test corrections; and peace and conflict studies examine how online narratives travel offline. What is missing is a theoretical account that connects what platforms select, how political actors exploit that selection, and what happens to trust between groups, and that specifies when the outcome is polarisation and violence rather than mere noise.

This paper develops such an account. It distinguishes three mechanisms: *amplification* (what ranking systems and networks select and deliver), *mobilisation* (how political entrepreneurs exploit that selection) and *erosion* (what happens to inter-group and institutional trust). Its central argument is that the principal harm of social media in plural societies is not misinformation as such but the *systematic over-supply of identity-threat content* to populations whose cohesion depends on its under-supply. Plural societies have historically sustained coexistence partly by keeping the most inflammatory claims about other groups out of shared public space; engagement-optimised systems reverse that filter.

The paper has three objectives: (i) to synthesise recent peer-reviewed evidence on social media, misinformation, hate speech and cohesion in Nigeria and other African states into an amplification–mobilisation–erosion model; (ii) to derive propositions about the conditions under which amplification produces polarisation and violence, and about the resilience factors (cross-cutting ties, trusted local intermediaries and institutional credibility) that moderate its effects; and (iii) to assess the limits of content moderation designed for other societies and draw implications for electoral management, peacebuilding and platform governance. It asks: Through what mechanisms does social media affect cohesion in plural African societies? Under what conditions do these mechanisms produce polarisation and violence? What moderates them, and what follows for governance?

2. LITERATURE REVIEW

2.1 CONCEPTUAL REVIEW

2.1.1 Misinformation, disinformation and information disorder

Misinformation refers to false or misleading content shared without intent to deceive, *disinformation* to such content shared deliberately, and *information disorder* to the wider ecosystem in which both circulate alongside rumour, propaganda and manipulated media. The Nigerian literature shows

© IIARJournal. All rights reserved.

how blurred these categories are in practice. During the 2023 campaign, candidates' teams engaged in prebunking, debunking, fake-news tagging and counter-disinformation while also introducing misleading material, so that information disorder is better understood as an ongoing exchange than as a one-way flow (Ibrahim & Ekdale, 2026). Generative AI adds synthetic media to the mix: in Kenya, deepfake clips reduced support for the depicted politician unless accompanied by sceptical comments (Wack & Prochaska, 2026).

2.1.2 Hate speech and identity-threat content

Hate speech refers to expression that attacks or demeans people on the basis of group identity. Analyses of Nigeria's 2023 presidential discourse identify derogation, stereotyping, dehumanisation, scapegoating, demonisation and polarisation as recurrent tactics across ethnic, religious, ideological and historical-alliance contexts (Alugbin & Adepoju, 2026). This paper uses the broader term *identity-threat content* for material, false or not, that frames an out-group as a threat to the in-group's security, status or survival. The #SayNoToRUGA episode illustrates the category: toxic discourse, ethnic profiling and misinformation combined to frame Fulani pastoralists as agents of "expansion and domination" (Ajetunmobi et al., 2024).

2.1.3 Algorithmic amplification

Amplification is the process by which some content reaches far more people than others. On feed-based platforms it is largely algorithmic, as ranking systems select for predicted engagement (Milli et al., 2025). On encrypted messaging platforms such as WhatsApp, which are central to African political communication, it is social, operating through forwarding within groups whose members trust one another (Fisher et al., 2024; Mudavadi et al., 2025). The model treats amplification as socio-technical: the combined selection performed by code and by networks.

2.1.4 Social cohesion and trust

Social cohesion refers to the quality of relations within and between groups in a society, including inter-group trust, a sense of shared belonging and confidence in common institutions. Survey evidence from 36 African countries finds that more frequent experiences of ethnic discrimination are associated with prioritising ethnic over national identity (Tuki, 2026a), and evidence from 39 countries finds that stronger ethnic than national attachment is associated with less support for interethnic marriage (Tuki, 2026b). Cohesion, in this sense, is a stock that can be built or depleted.

2.2 THEORETICAL REVIEW

Several theoretical lenses from recent African scholarship inform the model.

Social identity and common ingroup identity. Social identity theory explains why content that dramatises out-group threat is compelling: it activates group membership and its emotional stakes. Ademosu et al. (2026) use it to analyse ethnic and religious rhetoric in the 2023 Lagos governorship race, finding that charged wording provoked similarly hostile replies. The common ingroup identity model supplies the counterweight: among Yoruba adults in Nigeria, identification with a shared national identity predicted intentions to act for equality towards other groups (Cocco et al., 2024).

Ekeh's two publics. Wefwafwa et al. (2025) draw on Ekeh's concept of "two publics" (the ethnic, primordial public and the civic, national public) to explain how simultaneous ethnic and national belonging produce tensions, confrontations and eventually consensus in Kenyan WhatsApp groups. The lens helps explain why digital spaces can host both division and its negotiation.

Digital clientelism. Fisher et al. (2024) argue that pre-existing party institutionalisation shapes how African parties use WhatsApp, and that "digital clientelism" reproduces patronage structures online. In Nigeria's less institutionalised parties, informal, personality-based groups with weak message discipline create fertile conditions for misinformation. This lens explains the *mobilisation* mechanism.

Media systems dependency and securitisation. Media systems dependency theory, used by Mohammed and Adelakun (2023) to analyse reactions to Nigeria's 2021 Twitter ban, explains why abrupt state interventions provoke resistance among dependent users. Securitisation theory, used by Ndawana (2023) for Zimbabwe, explains why governments treat online protest as a security threat and thereby close opportunities for engagement. Together they inform the analysis of state responses.

© IIARJournal. All rights reserved.

Decolonial and post-digital perspectives on governance. Udupa et al. (2023) propose "ethical scaling" for content moderation, a transparent, community-participatory process grounded in a decolonial critique of the "human versus machine" framing. Hirblinger (2024) argues for a post-digital view of peacebuilding that does not separate digital tools from pre-existing practices and power relations. These lenses inform the analysis of moderation and peacebuilding.

2.3 EMPIRICAL REVIEW

2.3.1 Amplification: platforms and networks

Outside Africa, an audit of Twitter's engagement-based ranking found that it amplified angry, partisan and out-group-hostile content relative to a chronological feed, and that ranking by users' stated preferences reduced such content (Milli et al., 2025). In Africa, much amplification runs through messaging networks. In South Africa's 2024 elections, an analysis of 22,384 messages from 47 politically active WhatsApp groups found that fear-based appeals, identity-driven rhetoric and content imitating credible journalism were the most common tactics, with spread driven by emotional resonance and in-group trust (Gondwe, 2025). Kenyans interviewed about WhatsApp misinformation viewed sharing as a way of keeping others informed and hesitated to correct relatives for fear of embarrassment or conflict (Mudavadi et al., 2025). Users may also value such groups as "safe" spaces from state surveillance, even as their patterns of inclusion and exclusion threaten that safety (Ooko, 2023).

2.3.2 Mobilisation: political entrepreneurs

Nigerian evidence points to organised mobilisation. Politicians pay social media influencers, the so-called "Data Boys", to promote and defend their image covertly (Uwalaka et al., 2025). Network analysis of Nigerian Twitter identified a small number of influential accounts central to creating and spreading fake news (Gbaje et al., 2025). Party WhatsApp networks in Nigeria are informal and personality-based, with weak message discipline compared with Ghana's more institutionalised parties (Fisher et al., 2024). Mobilisation also has transnational dimensions: the Igbo diaspora sustains separatist sentiment in south-eastern Nigeria by reviving memories of the civil war (Onyemechalu & Ejiofor, 2024). Comparable dynamics appear elsewhere. In Ethiopia, political influencers and online "political armies" spread sectarian rhetoric on Facebook during the conflict in Amhara (Gessesse & Moges, 2026); in Kenya, the "hustlers versus dynasties" narrative spread online and contributed to polarisation and offline conflict (Ogenga & Gamez, 2026); and in Zimbabwe, parties used TikTok as a "politainment" tool aimed at young voters (Ureke, 2024).

2.3.3 Erosion: trust, polarisation and violence

In Nigeria, campaign hate speech is linked by voters to misinformation, fear-driven voting and low turnout (Aondover et al., 2026). The #SayNoToRUGA campaign polarised users along ethnic and regional lines (Ajetunmobi et al., 2024). In Ethiopia, language that alienated, dehumanised and mobilised people along ethnic lines deepened national fracture during the civil war (Mafu, 2024), and diaspora campaigns turned social media into a site of heated discursive conflict (Bitew, 2026). In Mali, Twitter both promoted and mediated conflict, and the boundary of who counts as a citizen journalist blurred (de Bruijn et al., 2025). In Kenyan electoral WhatsApp groups, gendered intimidation, often couched as humour, operates as both a means of dominance and a form of resistance, and can either support or limit deliberation (Wefwafwa, 2025). A review of the economics literature concludes that traditional and social media can affect violence, ethnic tensions, hate crime and protest (Petrova & Tapsoba, 2025).

2.3.4 Resilience and correction

Evidence on resilience is thinner but instructive. Ethiopian users responding to ethnic hate speech take on roles that include not only trolling but also peace-making, informing and guarding (Megersa & Minaye, 2023). In a Facebook Messenger experiment with roughly 15,000 users in Kenya and Nigeria, warning flags and related-article suggestions had no measurable effect, while accuracy prompts reduced misinformation sharing (Offer-Westort et al., 2024). In Kenya, "social truth queries" (peer replies that prompt readers to question a claim) reduced the perceived accuracy of election misinformation and trust in those who posted it (Wack & Jalbert, 2026), and sceptical comments partly offset the effect of deepfakes (Wack & Prochaska, 2026). Religious institutions in Kenya used social media to provide support during the pandemic (Birya & Mwamzandi,

© IIARJournal. All rights reserved.

2025), indicating the digital reach of trusted intermediaries. In Nigeria, Twitter users debunked false COVID-19 stories, while less-educated users who relied on social media and television were more likely to believe them (Uwalaka, 2023).

2.3.5 Moderation and state responses

Platform moderation performs poorly in African contexts. AI-driven moderation at Meta misreads local contexts in the Global South, both over-removing lawful speech and removing harmful content slowly, owing to limited investment, weak language coverage and power asymmetries (Balendra, 2025). Machine-learning tools struggle to keep pace with extreme speech in Kenya and elsewhere (Udupa et al., 2023), and researching online hate across African countries is itself constrained (McGeer & Stremlau, 2024). Fact-checkers outside Western democracies find falsehoods deeply tied to local context and see their work as supporting social cohesion (Vinhas & Bastos, 2025), but in Ethiopia and Mali they face abuse, fear of retaliation and dependence on funders (Badji et al., 2024; Azoulay, 2024). State responses have been blunt. When Nigeria suspended Twitter in 2021, most public reactions were negative and most surveyed students used VPNs to keep access (Mohammed & Adelakun, 2023). Nigeria's subsequent Code of Practice marks a shift towards co-regulation that places duties on platforms rather than users (Obia, 2025a), but platforms' symbolic power in Nigeria, which was evident in protesters' expectations of corporate support during #EndSARS, can make regulation harder (Serres, 2026; Obia, 2025b).

2.4 GAP IN THE LITERATURE

Three gaps follow. First, African studies of social media and cohesion rarely link platform selection, political mobilisation and trust erosion within one framework; each is studied separately. Second, the focus on falsity obscures the role of identity-threat content that need not be false. Third, there is little theory specifying the conditions under which online division becomes offline violence, and the resilience factors that prevent this. This paper addresses all three.

3. MATERIALS AND METHODS

3.1 RESEARCH DESIGN

The study is a theoretical, theory-building paper based on an integrative review of peer-reviewed literature; it collects no primary data. An integrative design suits a phenomenon spanning political communication, discourse studies, social psychology, peace and conflict studies and platform governance. The analytical logic is abductive: mechanisms reported across studies were compared, recurring patterns identified, and a model constructed that accounts for them and generates testable propositions.

3.2 POPULATION AND SAMPLE

The population comprised peer-reviewed journal articles published between 2023 and 2026 on social media, misinformation, disinformation, hate speech, political communication, ethnic identity, trust, conflict, fact-checking, content moderation and platform regulation, with priority given to Nigeria and other African states. Studies from outside Africa were admitted where they illuminated mechanisms, for example algorithmic ranking. The sampling technique was purposive, guided by the three mechanisms of the emerging model. The final sample comprised 45 articles (Table 1).

3.3 INCLUSION AND EXCLUSION CRITERIA

Studies were included if they (a) were peer-reviewed journal articles; (b) were published, in an issue or online ahead of print, between 2023 and 2026; (c) addressed one or more of the themes above; and (d) could be verified against the publisher's own record. Studies were excluded if they were book chapters, conference proceedings, preprints, reports, commentaries, theses or news items; if they were published before 2023; if their full author list or other bibliographic details could not be confirmed on the publisher page; or if they were outside scope. Foundational theory published before 2023 is discussed only as applied in the sampled literature.

3.4 DATA COLLECTION

Searches were conducted in October 2026 using scholarly web search restricted to the platforms of major academic publishers (including Sage, Oxford University Press, Cambridge University Press, Springer Nature, MDPI, Frontiers and Cogitatio).

© IIARJournal. All rights reserved.

Search strings combined place terms (*Nigeria*, *Kenya*, *Ethiopia*, *Ghana*, *South Africa*, *Africa*) with phenomenon terms (*misinformation*, *disinformation*, *fake news*, *hate speech*, *WhatsApp*, *Twitter/X*, *TikTok*, *influencers*, *deepfakes*, *algorithmic amplification*, *polarisation*, *ethnic identity*, *social cohesion*, *fact-checking*, *content moderation*, *platform regulation*, *peacebuilding*).

3.5 VERIFICATION AND BASES OF JUDGEMENT

Each candidate article was checked against its publisher page for title, authors, year, journal, volume, issue, pagination or article number and DOI, and the page was inspected for retraction notices, corrections or expressions of concern. None of the included articles carried such a notice. Of 57 candidate records that reached full-record verification, 45 met all criteria and 12 were excluded: the publisher page could not be accessed for confirmation (n = 6), the full author list could not be confirmed (n = 2), the item was a commentary (n = 1), it was published before 2023 (n = 1), or it was outside scope (n = 2). Earlier screening removed items that were evidently book chapters, reports or non-journal sources; these were not counted systematically, which is acknowledged as a limitation. Relevance was judged by whether a study reported evidence on amplification, mobilisation, erosion, resilience or governance.

Table. 1. Summary of source screening and verification

Stage / outcome	Records (n)
Candidate records reaching full-record verification	57
Excluded: publisher page not accessible for confirmation	6
Excluded: full author list not confirmable	2
Excluded: commentary, not research article	1
Excluded: published before 2023	1
Excluded: outside scope of review	2
Included in final sample	45

Note. All 45 included articles were confirmed against publisher records; none carried a retraction, correction or expression of concern at the time of checking (October 2026).

3.6 DATA ANALYSIS

Analysis proceeded in three stages. First, findings were extracted and coded against the three provisional mechanisms and against reported resilience factors and governance responses. Second, codes were compared across countries and platforms (feed-based platforms and encrypted messaging) to identify convergent and divergent patterns, summarised in an evidence map (Table 3). Third, recurrent mechanisms and conditions were formalised as propositions and integrated into a model (Figure 1). Throughout, claims were tested against contrary evidence, for example findings that WhatsApp groups also produce consensus (Wefwafwa et al., 2025) and that social media can support peacebuilding (Ndawana, 2023; Ogenga & Gamez, 2026).

3.7 VARIABLE MEASUREMENT

Although the paper is theoretical, its constructs are specified for future measurement. Table 2 presents proposed operational definitions and indicators.

Table. 2. Constructs, operational definitions and proposed indicators

Construct	Operational definition	Proposed indicators
Amplification	Reach of identity-threat content relative to other content, via ranking or forwarding	Share of identity-threat items in feeds vs. chronological baseline; forwarding depth in messaging groups; engagement per item
Mobilisation	Organised production and exploitation of identity-threat content by political actors	Volume of content from campaign, paid-influencer or diaspora accounts; network centrality of such accounts; timing relative to elections
Erosion	Decline in inter-group trust, shared identity and institutional confidence	Inter-group trust scales; ethnic vs. national identification; trust in electoral body; reported fear-driven voting
Resilience factors	Cross-cutting ties, trusted intermediaries, institutional credibility	Inter-group friendship and marriage; trust in religious/traditional leaders; perceived electoral-body independence
Outcome	Polarisation/violence vs. noise	Affective polarisation measures; incidence of election-related or communal violence; persistence of narratives offline

3.8 RELIABILITY, VALIDITY AND TRUSTWORTHINESS

Because the study is interpretive, rigour is addressed through trustworthiness criteria. *Credibility* was

© IIARJournal. All rights reserved.

supported by verifying every source against its publisher record and citing only what sources report. *Dependability* was supported by documenting the search terms, criteria and screening outcomes (Section 3.5, Table 1). *Confirmability* was supported by seeking contrary evidence. *Transferability* is bounded: the model is anchored in Nigerian and wider African evidence and should be tested before being applied elsewhere.

3.9 AI USAGE DISCLOSURE AND OTHER DISCLOSURES

Generative AI tools (a large language model) were used to assist with literature searching, checking bibliographic details against publisher records, and preparing an initial draft of the text and figure. All sources were checked against publisher records, and the author(s) reviewed, revised and take full responsibility for the content, interpretations and any errors. No AI tool is listed as an author. The study used only published literature; no human participants were involved and ethical approval was not required. All sources are listed in the references; no new datasets or code were generated.

4. RESULTS AND DISCUSSION

4.1 RESULTS

4.1.1 Evidence map

Table 3 maps the reviewed evidence onto the three mechanisms and the resilience factors. The entries are qualitative syntheses of the cited studies.

Table. 3. Evidence map: mechanisms, resilience and governance

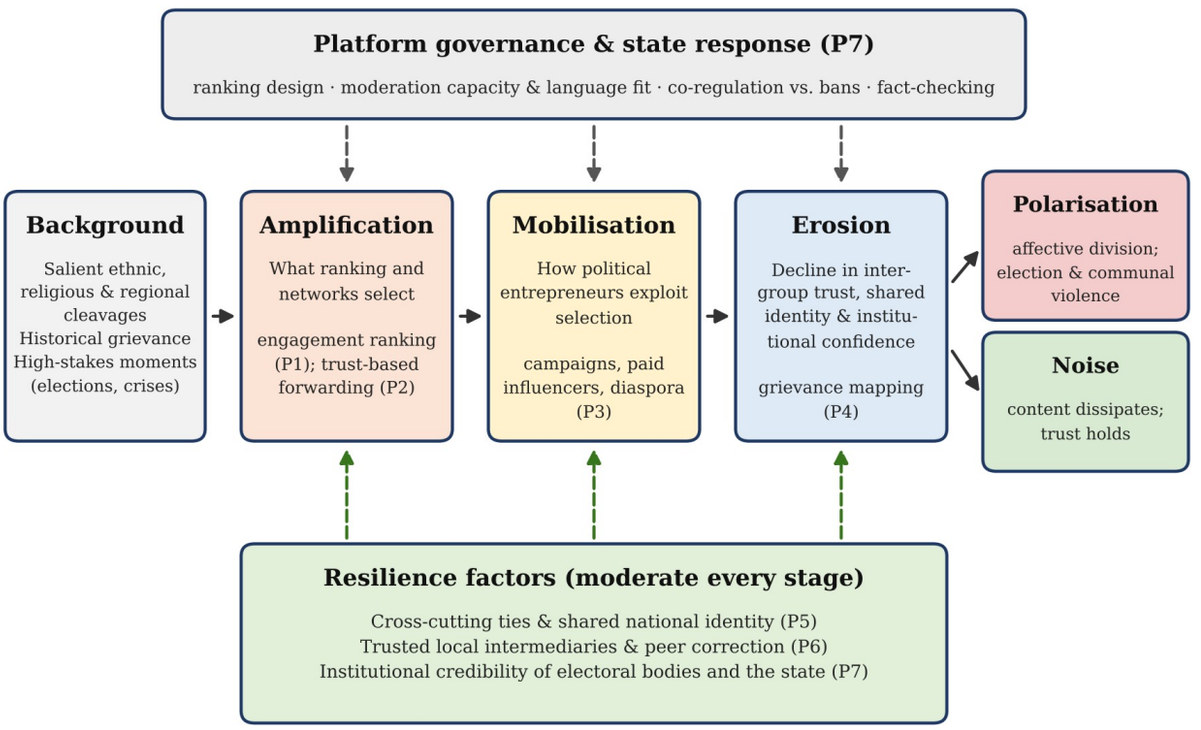
Component	Nigeria	Other African states	Implication
Amplification	Informal party WhatsApp networks (Fisher et al., 2024); Twitter as activism arena (Obia, 2025b)	Fear and identity appeals in WhatsApp groups (Gondwe, 2025); costly correction (Mudavadi et al., 2025)	Dual: algorithmic and trust-based

Note. Qualitative synthesis of the cited studies.

Three patterns stand out. First, amplification in Africa is dual: algorithmic on feed-based platforms and social on messaging platforms, where trust within groups substitutes for ranking and makes correction socially costly. Second, mobilisation is organised and resourced: paid influencers, small numbers of central accounts, personality-based party networks and diaspora entrepreneurs supply identity-threat content and exploit its amplification. Third, erosion is most visible at moments of high stakes (elections, separatist agitation and armed conflict), and resilience depends on actors and identities that cut across group lines.

© IIARJournal. All rights reserved.

Figure. 1. The amplification–mobilisation–erosion model of social media and cohesion in plural societies



Note. P1–P7 refer to the propositions in Section 4.1.3. Source: developed by the author(s) from the reviewed literature.

4.1.2 The amplification–mobilisation–erosion model

Figure 1 (above) presents the model. Background conditions (salient ethnic, religious and regional cleavages, histories of grievance and high-stakes political moments) shape the supply of identity-threat content. *Amplification* determines how much of that content reaches users, through engagement-based ranking and trust-based forwarding. *Mobilisation* describes political entrepreneurs (campaigns, paid influencers, diaspora networks and conflict actors) who produce and exploit amplified content. *Erosion* describes the cumulative effect on inter-group trust, shared identity and confidence in institutions. The outcome depends on resilience factors: where cross-cutting ties, trusted intermediaries and credible institutions are strong, amplified content tends to dissipate as noise; where they are weak, it feeds polarisation and, in some cases, violence. Platform governance and state responses act on every stage and can strengthen or weaken resilience.

4.1.3 Propositions

Proposition 1 (over-supply). Where ranking systems optimise for engagement, identity-threat content will be over-supplied relative to what users say they prefer, and the resulting harm will be greater in societies where cohesion depends on keeping such content out of shared space. This follows from audit evidence that engagement ranking amplifies out-group hostility beyond stated preferences (Milli et al., 2025) and from the prevalence of identity-threat frames in Nigerian discourse (Alugbin & Adepoju, 2026; Ajetunmobi et al., 2024).

Proposition 2 (closed networks). In encrypted messaging networks, amplification operates through in-group trust, so that identity-threat content spreads with social endorsement and corrections are resisted or withheld. This follows from evidence on emotional resonance and the dismissal of corrections in South Africa (Gondwe, 2025), the social costs of correction in Kenya (Mudavadi et al., 2025) and weak message discipline in Nigerian party networks (Fisher et al., 2024).

© IIARJournal. All rights reserved.

Proposition 3 (mobilisation at high-stakes moments). Amplification produces polarisation rather than noise when it coincides with organised mobilisation during high-stakes moments such as elections, security crises or separatist agitation. Nigerian hate speech peaks during political periods (Ridwanullah et al., 2025), campaigns and paid influencers supply misleading content (Ibrahim & Ekdale, 2026; Uwalaka et al., 2025), and voters link campaign hate speech to fear-driven choices and low turnout (Aondover et al., 2026).

Proposition 4 (grievance mapping). Online division escalates towards violence when amplified content maps onto historical grievances and existing patterns of conflict. Evidence from Ethiopia (Mafu, 2024; Gessesse & Moges, 2026), Mali (de Bruijn et al., 2025), Kenya (Ogenga & Gamez, 2026) and south-eastern Nigeria (Onyemechalu & Ejiofor, 2024) shows narratives that revive past wrongs and frame present threats, consistent with wider evidence that media can affect violence and ethnic tension (Petrova & Tapsoba, 2025).

Proposition 5 (cross-cutting ties and shared identity). The effects of amplification are weaker where cross-cutting ties and a shared national identity are strong. Common national identification predicts pro-equality intentions across Nigerian groups (Cocco et al., 2024), ethnic-over-national attachment is linked to discrimination and reduced openness to intermarriage (Tuki, 2026a, 2026b), and users who take peace-making and guarding roles counter hate speech (Megersa & Minaye, 2023).

Proposition 6 (trusted intermediaries and peer correction). Corrections are more effective when they come from trusted peers and local intermediaries than from platform labels. Peer-delivered social truth queries and sceptical comments reduced belief in misinformation and deepfakes in Kenya (Wack & Jalbert, 2026; Wack & Prochaska, 2026), accuracy prompts outperformed warning flags in Kenya and Nigeria (Offer-Westort et al., 2024), and religious institutions have digital reach (Birya & Mwamzandi, 2025).

Proposition 7 (institutional credibility and governance fit). The effects of amplification are greater where electoral and state institutions lack credibility, and where moderation is designed for other societies. Weak election administration in Nigeria (Oni et al., 2025) leaves space for contested narratives; AI moderation misreads Global South contexts (Balendra, 2025; Udupa et al., 2023); dominant fact-checking models neglect social and historical context (Vinhas & Bastos, 2025); and blunt state measures such as platform bans provoke circumvention rather than compliance (Mohammed & Adelakun, 2023).

4.1.4 The Nigerian anchoring case

Nigeria displays every element of the model. Its cleavages are salient and politicised, and identity-threat content is pervasive in campaign discourse (Alugbin & Adepoju, 2026; Ademosu et al., 2026). Amplification runs through both Twitter/X, which Nigerians treat as the foremost arena for activism and "dragging" (Obia, 2025b), and informal party WhatsApp networks (Fisher et al., 2024). Mobilisation is organised through paid influencers and central accounts (Uwalaka et al., 2025; Gbaje et al., 2025) and diaspora networks (Onyemechalu & Ejiofor, 2024). Resilience is uneven: shared national identity can promote pro-social action (Cocco et al., 2024), but electoral institutions face severe security pressures (Oni et al., 2025). Governance has moved from a platform ban (Mohammed & Adelakun, 2023) to co-regulation (Obia, 2025a), but platforms retain considerable symbolic power (Serres, 2026). In model terms, Nigeria combines high supply, dual amplification and organised mobilisation with uneven resilience, a configuration in which erosion is likely around elections and security crises.

4.2 DISCUSSION

The model reframes the problem from misinformation to over-supply. If the principal harm were falsity, fact-checking and labelling would be adequate remedies. The evidence suggests otherwise: warning flags had no measurable effect in Kenya and Nigeria (Offer-Westort et al., 2024), identity-threat content need not be false, and corrections in closed networks are often dismissed as propaganda (Gondwe, 2025). This is consistent with Vinhas and Bastos (2025), who argue that dominant fact-checking models treat misinformation as an individual behavioural problem and overlook its social and historical drivers.

The findings qualify social identity accounts. Identity-based rhetoric provokes reciprocal hostility (Ademosu et al., 2026), but the same digital spaces

© IIARJournal. All rights reserved.

can host negotiation and consensus (Wefwafwa et al., 2025), in line with the two-publics argument. The model locates the difference in resilience factors rather than in platforms alone. This also explains why digital peacebuilding is possible (Ndawana, 2023; Ogenga & Gamez, 2026) but cannot be separated from offline practices and power relations (Hirblinger, 2024).

The mobilisation mechanism extends the digital clientelism argument (Fisher et al., 2024). In Nigeria, the market for paid influence (Uwalaka et al., 2025) means that identity-threat content is not only spontaneously generated but commissioned, which makes political finance and campaign regulation part of the response. The erosion mechanism connects to evidence that discrimination pushes people towards ethnic over national identity (Tuki, 2026a): online over-supply and offline discrimination may reinforce one another.

For governance, the evidence supports three conclusions. First, moderation designed for other societies under-performs in Africa because of language, context and investment gaps (Balendra, 2025; Udupa et al., 2023); ethical scaling with community participation is a more promising model. Second, blunt state measures such as platform bans undermine legitimacy and are circumvented (Mohammed & Adelakun, 2023), whereas co-regulation that places duties of care on platforms (Obia, 2025a) aligns better with the model. Third, electoral management should treat information integrity as part of election security (Oni et al., 2025), including prebunking and rapid response mechanisms that draw on the campaign responses documented by Ibrahim and Ekdale (2026).

The paper has limitations. It is theoretical, and its propositions await testing. The sample was restricted to verified 2023–2026 journal articles, excluding foundational works and grey literature, and access constraints meant that some relevant articles could not be verified. Evidence on TikTok and on algorithmic ranking in African contexts remains thin, and much of the strongest experimental evidence comes from Kenya. Future research should test the propositions using platform data from Nigerian elections, comparative studies of WhatsApp and feed-based amplification, and field experiments with trusted intermediaries.

5. CONCLUSION AND RECOMMENDATION

This paper has argued that social media affect cohesion in plural African societies through three linked mechanisms (amplification, mobilisation and erosion) and that the principal harm is the systematic over-supply of identity-threat content to societies whose coexistence depends on its under-supply. Whether this over-supply produces polarisation and violence or dissipates as noise depends on the timing of political competition, the mapping of content onto historical grievance, and the strength of cross-cutting ties, trusted intermediaries and credible institutions. In Nigeria, high supply, dual amplification and organised mobilisation meet uneven resilience, making elections and security crises the moments of greatest risk.

Five recommendations follow. **First**, electoral management bodies, including Nigeria's Independent National Electoral Commission, should integrate information integrity into election security planning, with prebunking campaigns, rapid-response units and partnerships with fact-checkers and community leaders. **Second**, regulators should develop co-regulatory frameworks requiring platforms to report on ranking, moderation capacity in African languages and election-period risk assessments, rather than resorting to platform bans. **Third**, platforms should offer ranking based on stated preferences and reduce amplification of out-group hostility, especially during elections, and should invest in local-language moderation using participatory approaches. **Fourth**, peacebuilding actors and religious and traditional institutions should be supported to act as trusted intermediaries, including peer-correction programmes and digital literacy initiatives delivered through trusted networks. **Fifth**, political finance and campaign regulations should address paid influence operations by requiring disclosure of paid political promotion.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper makes four contributions. **Conceptually**, it shifts the focus from misinformation to identity-threat content and its over-supply, defining a category that captures harmful content regardless of falsity. **Theoretically**, it offers the amplification–mobilisation–erosion model, which integrates social identity, two-publics, digital clientelism,

© IIARJournal. All rights reserved.

securitisation and decolonial governance perspectives, and specifies the resilience factors that moderate harm. **Empirically**, it consolidates recent verified evidence from Nigeria and other African states into an evidence map, propositions and operational indicators for future research. **For policy**, it identifies electoral, regulatory, platform and peacebuilding instruments suited to plural African societies rather than imported from elsewhere.

7. 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

Ajetunmobi, U. O., Oseni, A. I., Onifade, A. B., & Adegboyega, T. A. (2024). Fear of expansion and domination: Toxic and ethnic frames amidst call to #SayNoToRUGA on Nigerian Twittersphere. *Media Watch, 15*(2), 246–266. https://doi.org/10.1177/09760911241230708

Alugbin, M., & Adepoju, O. T. (2026). Contexts and forms of hate speech in Nigeria's 2023 presidential election discourse. *Journal of Asian and African Studies, 61*(3), 2134–2149. https://doi.org/10.1177/00219096251332949

Aondover, E. M., Ademosu, I., & Acheme, R. O. (2026). Perception of hate speech in the 2023 presidential political campaigns on voter behaviour in Nigeria. *Frontiers in Political Science, 8,* Article 1750183. https://doi.org/10.3389/fpos.2026.1750183

Azoulay, L. (2024). Truth in the crossfire: The case of Ethiopia and fact-checking in authoritarian contexts. *Media and Communication, 12,* Article 8785. https://doi.org/10.17645/mac.8785

Badji, S. D., Orgeret, K. S., & Mutsvairo, B. (2024). An exploratory study of fact-checking practices in conflict and authoritarian contexts. *Media and Communication, 12,* Article 8698. https://doi.org/10.17645/mac.8698

Balendra, S. (2025). Meta's AI moderation and free speech: Ongoing challenges in the Global South. *Cambridge Forum on AI: Law and Governance, 1,* Article e21. https://doi.org/10.1017/cfl.2025.5

Birya, R. K., & Mwamzandi, M. (2025). Religious institutions' use of social media platforms during COVID-19 in Kenya. *Social Media + Society, 11*(3). https://doi.org/10.1177/20563051251372530

Bitew, B. (2026). Hands-off Ethiopia: Western interventions and netizen reactions in the Ethiopian civil war, 2020–2022. *Insight on Africa, 18*(1), 49–66. https://doi.org/10.1177/09750878241312187

Cocco, V. M., Vezzali, L., Kola-Daisi, T. I., & Çakal, H. (2024). The role of common ingroup identity in promoting social change among tribes in Nigeria. *Group Processes & Intergroup Relations, 27*(1), 142–157. https://doi.org/10.1177/13684302231162038

de Bruijn, M., Mutsvairo, B., Bruls, L., Cissé, M. G., Langguth, J., Orgeret, K. S., Badji, S. D., Moges, M. A., & Schroeder, D. T. (2025). "Aren't we all journalists?" Citizen journalism, disinformation and the weaponization of social media in conflict torn Mali. *Journalism, 26*(5), 1027–1046. https://doi.org/10.1177/14648849241312743

Fisher, J., Gadjanova, E., & Hitchen, J. (2024). WhatsApp and political communication in West Africa: Accounting for differences in parties' organization and message discipline online. *Party Politics, 30*(5), 934–948. https://doi.org/10.1177/13540688231188690

Gbaje, E. S., Agwu, C., Odigie, I. O., & Yani, S. D. (2025). Curtailing fake news creation and dissemination in Nigeria: Twitter social network and sentiment analysis approaches. *Journal of Information Science, 51*(4), 987–999. https://doi.org/10.1177/01655515231160029

Gessesse, K. B., & Moges, M. A. (2026). Social media discourses amidst ethnopolitical

© IIARJournal. All rights reserved.

extremism and conflict: The case of Ethiopia. *Discourse Studies, 28*(3), 408–433. https://doi.org/10.1177/14614456251374251

Gondwe, G. (2025). AI-driven disinformation and political influence on WhatsApp in South Africa's 2024 elections. *The International Journal of Press/Politics*. Advance online publication. https://doi.org/10.1177/19401612251395434

Hirblinger, A. T. (2024). When the digits don't add up: Research strategies for post-digital peacebuilding. *Cooperation and Conflict, 59*(3). https://doi.org/10.1177/00108367231184727

Ibrahim, A. A., & Ekdale, B. (2026). Political campaign responses to information disorder: A case study of the 2023 Nigerian presidential elections. *The International Journal of Press/Politics, 31*(3), 595–615. https://doi.org/10.1177/19401612251324985

Mafu, L. (2024). Discourse structures, weaponization of language and Ethiopia's civil war. *SAGE Open, 14*(3). https://doi.org/10.1177/21582440241262755

McGeer, C., & Stremlau, N. (2024). Researching hate speech online: Exploring the potential and limitations of Facebook as a survey tool in Africa. *Methodological Innovations, 17*(3), 172–186. https://doi.org/10.1177/20597991241264838

Megersa, T., & Minaye, A. (2023). Social media users' online behavior with regard to the circulation of hate speech. *Frontiers in Communication, 8*, Article 1276245. https://doi.org/10.3389/fcomm.2023.1276245

Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus, 4*(3), Article pgaf062. https://doi.org/10.1093/pnasnexus/pgaf062

Mohammed, A., & Adelakun, L. A. (2023). The 2021 Nigerian Twitter ban: A text-analytics and survey insight into public reactions and outcomes in the early weeks of the ban. *Communication and the Public, 8*(4), 390–401. https://doi.org/10.1177/20570473231209077

Mudavadi, K. C., Tully, M., & Lomoywara, D. B. (2025). Exploring Kenyans' interactions with misinformation on WhatsApp. *Mobile Media & Communication, 13*(1), 69–87. https://doi.org/10.1177/20501579241269651

Ndawana, E. (2023). The potential role of social media in peacebuilding in Zimbabwe. *Journal of Peacebuilding & Development, 18*(1), 53–67. https://doi.org/10.1177/15423166231154408

Obia, V. (2025a). 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

Obia, V. (2025b). Twitter activism: Understanding the Twittersphere as the foremost community for activism and dragging in Nigeria. *New Media & Society, 27*(1), 151–167. https://doi.org/10.1177/14614448231172967

Offer-Westort, M., Rosenzweig, L. R., & Athey, S. (2024). Battling the coronavirus 'infodemic' among social media users in Kenya and Nigeria. *Nature Human Behaviour, 8*(5), 823–834. https://doi.org/10.1038/s41562-023-01810-7

Ogenga, F., & Gamez, S. (2026). "Hustlers versus dynasties": Social media narratives, the 2022 presidential election in Kenya and digital peacebuilding. *Journal of Peacebuilding & Development, 21*(1), 123–139. https://doi.org/10.1177/15423166251379568

Oni, E. O., Faluyi, O. T., Asumu, L. O., & Olutola, A. A. (2025). Security challenges and election administration in Nigeria's fourth republic. *Frontiers in Political Science, 7*, Article 1458303. https://doi.org/10.3389/fpos.2025.1458303

Onyemechalu, S. J., & Ejiofor, P. F. (2024). Long-distance nationalism, diaspora mobilisation, and the struggle for Biafran self-determination in Nigeria. *Ethnicities, 24*(3). https://doi.org/10.1177/14687968231208307

© *IIARJournal. All rights reserved.*

Ooko, G. A. (2023). In pursuit of a “safe” space for political participation: A study of selected WhatsApp communities in Kenya. *Journalism and Media, 4*(2), 506–529. https://doi.org/10.3390/journalmedia4020032

Petrova, M., & Tapsoba, A. (2025). Information and conflict: From the role of (social) media and public opinion to big data and forecasting. *Economic Policy, 40*(124), 845–878. https://doi.org/10.1093/epolic/eiaf012

Ridwanullah, A. O., Sule, S. Y., Usman, B., & Abdulsalam, L. U. (2025). Politicization of hate and weaponization of Twitter/X in a polarized digital space in Nigeria. *Journal of Asian and African Studies, 60*(5), 3350–3370. https://doi.org/10.1177/00219096241230500

Serres, J. (2026). Political protest seeks corporate sponsor: Rethinking Big Tech, youth agency, and the reordering of power from Nigeria. *Convergence: The International Journal of Research into New Media Technologies*. Advance online publication. https://doi.org/10.1177/13548565261475283

Tuki, D. (2026a). Ethnic discrimination and the dynamics of belonging in Africa. *Political Studies Review*. Advance online publication. https://doi.org/10.1177/14789299261458205

Tuki, D. (2026b). Hearts and boundaries: Identity and attitudes toward interethnic marriage in Africa. *Journal of Race, Ethnicity, and Politics*. Advance online publication. https://doi.org/10.1017/rep.2026.10080

Udupa, S., Maronikolakis, A., & Wisiorek, A. (2023). Ethical scaling for content moderation: Extreme speech and the (in)significance of artificial intelligence. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517231172424

Ureke, O. (2024). Politics at play: TikTok and digital persuasion in Zimbabwe’s 2023 general elections. *Africa Spectrum, 59*(2). https://doi.org/10.1177/00020397241241335

Uwalaka, T. (2023). ‘Abba Kyari did not die of coronavirus’: Social media and fake news during a global pandemic in Nigeria. *Media International Australia, 188*(1), 18–33. https://doi.org/10.1177/1329878X221101216

Uwalaka, T., Amadi, F., & Enyindah, S. C. (2025). Social media influencers and political influence operations: The Data Boys example in Nigeria. *Public Relations Inquiry, 14*(1), 89–115. https://doi.org/10.1177/2046147X241301524

Vinhas, O., & Bastos, M. (2025). When fact-checking is not WEIRD: Negotiating consensus outside Western, educated, industrialized, rich, and democratic countries. *The International Journal of Press/Politics, 30*(1), 256–276. https://doi.org/10.1177/19401612231221801

Wack, M., & Jalbert, M. (2026). Social truth queries as a novel method for combating misinformation: Evidence from Kenya. *The International Journal of Press/Politics, 31*(3). https://doi.org/10.1177/19401612241301671

Wack, M., & Prochaska, S. (2026). Making sense of AI-generated disinformation: How audience interpretations influence the impact of deepfakes in Kenya. *Social Media + Society, 12*(3). https://doi.org/10.1177/20563051261462092

Wefwafwa, J. A. (2025). Vulgarities, vulnerabilities, and vengeance: Gendered intimidation and WhatsApp use in electoral politics in Kenya. *Global Studies Quarterly, 5*(3), Article ksaf065. https://doi.org/10.1093/isagsq/ksaf065

Wefwafwa, J. A., Wekesa, B., & Gagliardone, I. (2025). Tensions, confrontations, and consensus: WhatsApp use in Kenyan electoral politics. *Social Media + Society, 11*(1). https://doi.org/10.1177/20563051251315252

© IIARJournal. All rights reserved.