IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Security, Peace, and Conflict Studies

Vol 1, No 1, 2026

Digital Extremism and Counter-Radicalisation: A Theoretical Account of Algorithmic Pathways, Online Mobilisation and the Limits of Content Moderation in African Contexts

NWOGBOJI Emmanuel Usulor[1], Igwe Emmanuel Sunday[2]

[1,2] Veristzon Global Research Lab

ABSTRACT

Violent extremist movements in Africa recruit, coordinate and propagandise through the same platforms and messaging applications that carry ordinary social life, and the algorithms that sort content for engagement are widely suspected of drawing individuals toward them. This theoretical paper develops an account of digital extremism and its counters in African contexts. Drawing on radicalisation theory and the terrorism-studies literature, on recent audit and experimental research on recommendation algorithms, on the scholarship on deplatforming and platform migration, and on research on content moderation in the Global South and in low-resource languages, it distinguishes three functions digital platforms perform for extremist movements, exposure (bringing content to the susceptible), community (connecting them to one another) and operation (coordinating action), and argues that recommendation algorithms contribute most to the first, moderately to the second and little to the third, which has migrated to encrypted and fringe channels beyond moderation. It shows that content moderation designed in and for the Global North fails in African contexts for reasons of language, context, platform investment and the political ambiguity of what counts as extremism, and that the movements most lethal on the continent mobilise primarily through territorial governance and offline networks rather than through feeds. The paper derives seven propositions about where online intervention can reduce mobilisation and where it cannot, and about the risk that counter-extremism measures expand surveillance and suppress dissent. Nigeria's experience with Boko Haram and the Islamic State West Africa Province anchors the analysis. The paper closes with implications for governments, platforms, civil society and research.

Keywords: radicalisation; violent extremism; online mobilisation; recommendation algorithms; content moderation

Digital Extremism and Counter-Radicalisation: A Theoretical Account of Algorithmic Pathways, Online Mobilisation and the Limits of Content Moderation in African Contexts

1. INTRODUCTION

The image that dominates public discussion of digital extremism is the rabbit hole: a young person begins watching ordinary content, is led by a recommendation algorithm toward progressively more extreme material, and emerges radicalised. The image has driven policy across the world, from demands that platforms change their algorithms to legal duties to remove terrorist content within hours. It has also shaped how African governments and their partners think about the threat posed by movements such as Boko Haram, the Islamic State West Africa Province (ISWAP), al-Shabaab and the jihadist coalitions of the central Sahel. If radicalisation happens in the feed, counter-radicalisation should happen in the feed too.

This paper argues that the image is partly right and largely misplaced, especially in Africa. It is partly right because recommendation systems do shape exposure, and audit research has found that they can steer users toward congenial and, at the margins, problematic content (Haroon et al., 2023; Shin & Jitkajornwanich, 2024). It is largely misplaced because experimental evidence suggests that short-term algorithmic exposure has limited effects on attitudes (Liu et al., 2025), because the organisational work of extremist movements has migrated to encrypted and fringe channels that moderation does not reach (Jost & Dogruel, 2023; McMinimy et al., 2023; Weimann & Pack, 2026), and because the movements that kill most people on the African continent mobilise primarily through territorial control, local grievance and offline networks rather than through algorithmic feeds (Stoddard, 2023). A counter-radicalisation strategy built on the rabbit-hole image will therefore act on the part of the problem it can see while missing the part that matters most.

The paper develops this argument through a functional decomposition. Digital platforms perform three functions for extremist movements: exposure, community and operation. Each function relies on different platform features, is affected by different interventions and has a different weight in the radicalisation process. The paper then examines why content moderation, the principal online intervention, performs poorly in African contexts, and what the consequences of that failure, and of the measures governments take in response to it, are for security and for civil liberties. It derives propositions that can guide research and policy and anchors the analysis in Nigeria.

The analysis begins with the contested evidence on online radicalisation and then develops the exposure–community–operation decomposition, before asking where African extremist movements actually mobilise. It examines why content moderation performs poorly in African contexts and turns to the counter-extremism state and its unintended effects, from which a set of propositions follows. It closes by anchoring the argument in Nigeria's war with Boko Haram and ISWAP and drawing out the implications for governments, platforms and civil society.

2. LITERATURE REVIEW

2.1 Online Radicalisation: The Contested Evidence

A quarter-century of research

Research on the internet and radicalisation is now a quarter-century old. Rothut et al. (2026), reviewing twenty-five years of online-radicalisation research, find that the field has moved from treating the internet as a separate space of radicalisation toward understanding pathways as hybrid, combining online and offline exposure, relationships and action. The shift matters because it relocates the question from whether online content radicalises to how online and offline processes interact. Berjawi et al. (2023), surveying computational approaches to detecting and preventing online radicalisation, find a large body of technical work on detection and a much thinner body on prevention, with persistent gaps in evaluation, in multilingual coverage and in the connection between detected signals and real-world outcomes.

The algorithmic question

The specific question of whether recommendation algorithms radicalise has produced contested evidence. Haroon et al. (2023), auditing YouTube's recommendation system with a large population of automated accounts, find that the algorithm recommends ideologically congenial content to partisan users and that congeniality increases deeper in the recommendation trail for some users; they do not find meaningful increases in ideological extremity deeper in the trail, but they do find a growing, though small, share of recommendations

© IIARJournal. All rights reserved.

from channels categorised as problematic. Shin and Jitkajornwanich (2024), auditing TikTok through a reverse-engineering method, argue that its recommendation algorithm can play a role in self-radicalisation by narrowing the content users see. Against these, Liu et al. (2025), in naturalistic experiments with nearly nine thousand participants on a YouTube-like platform, manipulate recommendations to simulate filter bubbles and rabbit holes and find limited effects on opinions from short-term exposure. The evidence, taken together, supports a modest claim: algorithms shape what people are exposed to, can concentrate exposure for those already inclined, and contribute at the margins to contact with extreme content; they do not, on current evidence, convert the uninclined.

Platform migration

A second body of research concerns where extremist activity goes when mainstream platforms act against it. McMinimy et al. (2023) examine the influence of online restriction on the official media products of the Islamic State, showing how deplatforming efforts reshaped the production and distribution of the group's media. Jost and Dogruel (2023) analyse radical mobilisation on Telegram channels during the pandemic, in the wake of deplatforming that pushed radical actors from mainstream platforms. Weimann and Pack (2026) trace the migration of extremist content to successive alternative platforms, most recently TamTam, documenting the cat-and-mouse dynamic in which removal from one platform produces relocation to another. Whittaker and Craanen (2025) examine the unintended consequences of terrorist-content removal through the case of a fringe video platform, and Lakomy (2023) asks why online countering-violent-extremism strategies do not work against digital jihad, pointing to the resilience and adaptability of jihadist online ecosystems.

What the evidence implies

The evidence implies that mainstream-platform interventions, algorithmic change, content removal, account suspension, act on a shrinking and less operationally important part of the extremist online presence. They can reduce casual exposure on mainstream platforms. They displace rather than eliminate committed activity. And they have little effect on the encrypted channels where coordination now takes place.

2.2 Three Functions: Exposure, Community, Operation

Digital platforms perform three analytically distinct functions for extremist movements. The radicalisation literature has tended to treat them together; separating them clarifies what online intervention can and cannot do.

Exposure

Exposure is the function of bringing extremist content to people who were not seeking it. It is the function most dependent on recommendation algorithms and on the open, searchable architecture of mainstream platforms. It is also the function on which audit research has concentrated (Haroon et al., 2023; Shin & Jitkajornwanich, 2024). The evidence reviewed above suggests that algorithms contribute to exposure but that exposure alone has limited attitudinal effect (Liu et al., 2025). Exposure is necessary for radicalisation only in the trivial sense that a person must encounter an idea to adopt it; the encounter can occur through many channels, most of them offline in contexts where internet penetration is low and where extremist movements are physically present.

Community

Community is the function of connecting individuals who have encountered extremist ideas to others who share them, providing belonging, identity, validation and the social pressure that sustains commitment. The radicalisation literature consistently identifies social relationships as more important than content in moving people from sympathy to involvement (Rothut et al., 2026). Community online is formed in groups, channels, comment threads and messaging networks, some on mainstream platforms and increasingly on Telegram and similar services (Jost & Dogruel, 2023). Recommendation algorithms contribute to community formation by suggesting groups and accounts, but much community formation is driven by users' own choices and by invitation through existing networks.

Operation

Operation is the function of coordinating action: planning attacks, moving money, recruiting fighters, distributing instructions, and managing the logistics of an armed movement. It is the function of greatest security significance and the one least affected by mainstream-platform intervention. Operational communication has moved to encrypted messaging and to fringe platforms precisely because mainstream platforms remove it (McMinimy et al., 2023; Weimann & Pack, 2026). Content moderation

© IIARJournal. All rights reserved.

does not reach encrypted channels by design, and recommendation algorithms play no role in them.

Weighting the functions

The decomposition yields an inverse relationship. Recommendation algorithms matter most for exposure, which matters least for violence; they matter little for operation, which matters most. Mainstream-platform intervention is therefore best suited to reducing casual exposure and least suited to disrupting the activity that produces attacks. This does not make exposure unimportant, propaganda shapes the narratives through which movements recruit and justify violence, as Osisanwo's (2025) analysis of the discursive strategies in Boko Haram statements shows, but it locates the platform's contribution in the narrative environment rather than in the mechanics of mobilisation.

2.3 Where African Extremist Movements Mobilise

The decomposition developed above was derived largely from research on Western platforms and audiences. Applied to the movements responsible for most extremist violence in Africa, it must be weighted differently, because those movements mobilise in conditions that the online-radicalisation literature does not describe.

Territorial governance

The first condition is territory. Several of the continent's most lethal movements control or contest territory and the populations within it. Stoddard (2023) shows that ISWAP has pursued a strategy of competitive control in the Lake Chad region, seeking the compliance and support of local populations by offering order, dispute resolution, regulated taxation and protection, in deliberate contrast to the indiscriminate violence of its rival faction and to the conduct of state forces. A movement that governs recruits through the relationships that governance creates: those who trade, farm and fish under its authority, who bring disputes to its courts, and who depend on its protection. Online propaganda in this setting is an accompaniment to mobilisation rather than its engine. The rabbit hole is replaced by the market, the mosque and the checkpoint.

Grievance and the conduct of the state

The second condition is grievance, and in particular grievance against the state. The radicalisation literature has long recognised that experiences of injustice, humiliation and exclusion matter more than exposure to ideas in moving people toward violent movements (Rothut et al., 2026). In northern Nigeria and the Sahel, those experiences are frequently produced by the state's own counter-insurgency: mass arrests, detention without trial, killings of civilians and the destruction of livelihoods. Osisanwo (2025) shows that Boko Haram's statements deployed discursive strategies designed to authenticate the movement's violence and position it as a response to injustice; the strategy works to the extent that the injustice is experienced. Grievance of this kind is produced offline and is answered, if at all, offline.

Economic, kinship and coercive pathways

The third condition is the variety of non-ideological pathways into armed movements. Ojewale (2025) shows that in Nigeria's North-West, conflict, terrorism and banditry have become intertwined, with armed groups drawing members through economic opportunity, communal conflict, kinship networks and the protection they offer in a context where the state offers little. Forced recruitment and abduction, including of children, have supplied fighters to Boko Haram factions. None of these pathways runs through recommendation algorithms, and none is affected by content removal.

Connectivity

The fourth condition is connectivity itself. The rural areas in which these movements recruit most heavily have low internet penetration, intermittent mobile coverage, and populations for whom the radio, the phone call and the voice note are more important than the feed. Where the online sphere matters, it matters through messaging applications and voice rather than through the algorithmic platforms on which counter-extremism policy concentrates. The online audience for extremist propaganda is disproportionately urban and diasporic, and it is not the audience from which most fighters are drawn.

Implication

The implication is that the weighting of platform functions developed above must be adjusted further for African contexts. Exposure through algorithms matters even less, because the populations most at risk are least exposed to algorithmic feeds. Community is formed primarily offline, through the social relations of territory, kinship and shared grievance. Operation runs through mobile phones and messaging applications that no moderation regime reaches. Online counter-extremism, in these conditions, addresses the periphery of mobilisation.

© IIARJournal. All rights reserved.

2.4 Content Moderation in African Contexts

Content moderation is the principal instrument of online counter-extremism, combining automated classifiers with human review against platform policies and, increasingly, legal obligations. Its performance in African contexts is poor for four reasons.

Language

Most African languages are low-resource in the sense that natural-language-processing tools are trained on little data in them. Nicholas and Bhatia (2023) analyse the problem of automated content moderation in low-resource languages and show that classifiers perform worse, that multilingual models transfer poorly, and that platforms have incentives to underinvest in moderating languages with small commercial markets. Ayalew (2025a), examining Facebook's moderation in Ethiopia, documents linguistic blind spots that leave harmful content in several Ethiopian languages unmoderated while lawful speech is removed. In northern Nigeria and the Sahel, extremist content circulates in Hausa, Kanuri, Fulfulde, Arabic and other languages for which moderation capacity is thin, and in audio and video formats that text classifiers do not process.

Context

Extremist content is defined by context as much as by words. A sermon, a song, a battlefield video or a statement of grievance may be propaganda in one setting and news, worship or political speech in another. Balendra (2025) shows that AI-driven moderation in the Global South produces both over-removal, the censorship of lawful content, and slow removal of harmful material, because classifiers built on Western-centric frameworks misread local contexts. In African settings where religious and political speech are entangled and where insurgent and state narratives compete, the contextual judgement required to distinguish extremist propaganda from legitimate expression is beyond automated systems and frequently beyond moderators without local knowledge.

Investment

Platform investment in moderation follows commercial value and regulatory pressure, both of which are lower in most African markets than in Europe and North America. Nicholas and Bhatia (2023) and Ayalew (2025a) identify underinvestment as a structural cause of moderation failure. The consequence is that the populations most exposed to extremist violence are served by the least-resourced moderation systems.

Political ambiguity

The fourth reason is the hardest. What counts as extremism is politically contested, and in many African states governments designate opposition movements, separatist groups, protest movements and critical journalism as extremist or terrorist. Ayalew (2025b), offering a third-world critique of the human-rights-based approach to content moderation, argues that the international framework assumes a neutral state and a rights-respecting regulatory environment that do not obtain in many African contexts, and that platforms applying it may either defer to governments that abuse it or override local legal orders in ways that lack legitimacy. A platform asked by an African government to remove 'extremist' content faces the problem that compliance may suppress dissent and refusal may leave genuine threats in place.

The result

The result is a moderation regime that is simultaneously too weak and too strong: too weak to remove genuine extremist content in local languages and contexts, and too strong in its removal of lawful speech misread by classifiers or targeted by governments. Neither failure is corrected by more aggressive moderation, which increases over-removal, or by less, which increases exposure.

2.5 The Counter-Extremism State

Securitisation of the online sphere

Where platforms fail, governments act. The repertoire includes surveillance of social media and messaging, legal duties on platforms to remove content, criminalisation of online speech under terrorism and cybercrime laws, internet shutdowns in conflict areas, and the designation of groups and individuals as extremist. Each measure is justified by the threat of digital extremism, and each can be turned against civil society, journalists and political opposition. Njoku (2025), reviewing the literature on counter-terrorism measures and civil society organisations, documents how such measures constrain the space for civil society, including in Nigeria, and calls for research on the securitisation of civic life. Khaemba (2026) notes that preventing and countering violent extremism in sub-Saharan Africa has been dominated by military and securitised approaches and argues for the role of non-state, community-rooted actors.

© IIARJournal. All rights reserved.

The feedback loop

Securitised responses to online extremism can feed the grievances on which extremist movements draw. Surveillance and prosecution of online speech fall disproportionately on communities already marginalised; shutdowns cut off information and livelihoods in conflict areas; heavy-handed designation alienates populations whose cooperation counter-insurgency needs. Lakomy (2023) observes that online counter-extremism can produce effects contrary to its aims, and Whittaker and Craanen (2025) document unintended consequences of removal. In African contexts where state violence is itself a driver of recruitment, a counter-extremism state that expands surveillance and repression in the name of fighting online radicalisation may increase the radicalisation it targets.

The alternative

The alternative is not inaction but a different allocation of effort: from removal toward counter-narrative and credible voices in local languages; from surveillance toward community-based prevention; from platform-level intervention toward the offline drivers, grievance, governance failure, state abuse, that the radicalisation literature identifies. Khaemba (2026) and Njoku (2025) point to civil society and social enterprise as actors better placed than the security state for much of this work, provided they are not themselves constrained by counter-terrorism regulation.

3. CONCEPTUAL FRAMEWORK

Proposition 1. Recommendation algorithms contribute to exposure to extremist content but have limited independent effect on radicalisation, so that algorithmic interventions will reduce casual exposure on mainstream platforms without measurably reducing recruitment into violent movements (Haroon et al., 2023; Liu et al., 2025). Proposition 2. Mainstream-platform removal of extremist content displaces committed activity to encrypted and fringe channels rather than eliminating it, and the displacement concentrates the most operationally significant activity where intervention is least possible (McMinimy et al., 2023; Weimann & Pack, 2026). Proposition 3. Automated content moderation in low-resource African languages will produce higher rates of both false negatives and false positives than in high-resource languages, and the error rates will be highest for audio and video content in local languages (Nicholas & Bhatia, 2023; Ayalew, 2025a). Proposition 4. Where governments designate opposition, separatist or protest movements as extremist, platform compliance with removal requests will suppress legitimate expression in proportion to the breadth of the designation, and moderation will become an instrument of political control (Ayalew, 2025b; Balendra, 2025). Proposition 5. For movements that control territory and provide governance, online propaganda is a secondary recruitment channel, and online counter-extremism will have smaller effects than interventions addressing governance, grievance and state conduct in the areas they control (Stoddard, 2023). Proposition 6. Securitised responses to online extremism that expand surveillance, criminalise speech or shut down connectivity will increase grievance among targeted communities, and the net effect on radicalisation will be positive where state abuse is already a recruitment driver (Njoku, 2025; Lakomy, 2023). Proposition 7. Counter-narrative and prevention work delivered by credible local actors in local languages will outperform platform removal in reducing exposure effects, but only where those actors are not constrained by counter-terrorism regulation (Khaemba, 2026).

The propositions are stated as conditional claims that comparative and experimental research can test, and that policy can use as design constraints.

4. ANALYSIS AND DISCUSSION

4.1 Nigeria: Boko Haram, ISWAP and the Online Question

The movements

Boko Haram and its splinter ISWAP have conducted one of the deadliest insurgencies on the continent for more than fifteen years. Both have made extensive use of media: Boko Haram's leaders issued video and audio statements whose discursive strategies Osisanwo (2025) analyses, showing how they authenticated violence and invited audiences to the movement's side. ISWAP, affiliated with the Islamic State, has benefited from the latter's media apparatus. Yet the most consequential feature of ISWAP's strategy, as Stoddard (2023) shows, is territorial: it has competed for control of populations in the Lake Chad region by providing order, adjudication, taxation and limited services, seeking hearts and minds through governance rather than through feeds. Recruitment into both movements has been shaped by local economies of violence, forced conscription, kinship networks and

© IIARJournal. All rights reserved.

the conduct of state forces as much as by ideology encountered online.

Exposure

Proposition 1 and Proposition 5 apply. The exposure function of digital platforms matters for Nigerian audiences in cities and among the diaspora, where propaganda shapes perceptions of the conflict, but the populations from which the movements recruit most heavily in the north-east are those with the lowest internet penetration and the strongest physical exposure to the movements themselves. Algorithmic intervention on mainstream platforms will reach the former and not the latter.

Moderation

Proposition 3 applies with force. Content relevant to the insurgency circulates in Hausa, Kanuri, Arabic and other languages, often as audio sermons and video, and moderation capacity in these languages and formats is thin. At the same time Proposition 4 is relevant: Nigeria has used terrorism designations against movements beyond the jihadist insurgency, and has suspended a major platform nationally, so that platform compliance with government requests carries risks for legitimate political expression.

The counter-extremism state

Proposition 6 is the most important for Nigeria. The conduct of security forces in the north-east, mass detention, extrajudicial killings, and civilian harm from operations including the airstrikes and drone strikes that have become central to counter-insurgency in the region (Okpaleke et al., 2023), has been a documented driver of grievance. Njoku (2025) shows that counter-terrorism measures have constrained Nigerian civil society. A counter-extremism strategy that expands online surveillance and criminalisation while leaving state conduct unaddressed acts on the smallest cause of recruitment and risks strengthening the largest. The banditry crisis in the North-West, which Ojewale (2025) shows has intertwined with jihadist activity, adds a further dimension: armed groups there use mobile phones for ransom negotiation and coordination, an operational function that platform moderation cannot touch.

What would help

The framework suggests three priorities for Nigeria. First, invest in counter-narrative and prevention capacity in Hausa, Kanuri, Fulfulde and other languages through credible local actors, religious leaders, community organisations, returnees, while protecting those actors from counter-terrorism constraints (Khaemba, 2026). Second, press platforms for moderation capacity in Nigerian languages and audio formats, with transparency about government requests. Third, treat state conduct and governance in the north-east and north-west as the primary counter-radicalisation agenda, recognising that ISWAP's competitive governance (Stoddard, 2023) is countered by better governance rather than by better moderation.

4.2 Implications

For governments

African governments should weight counter-radicalisation effort toward the offline drivers of recruitment and toward prevention by credible local actors, and away from surveillance, criminalisation and shutdowns whose net effect may be to increase grievance. Where online measures are used, they should be narrowly targeted at operational content, subject to judicial oversight, and transparent.

For platforms

Platforms should invest in moderation capacity in African languages and audio-visual formats in proportion to the harm at stake rather than the market value, publish data on government removal requests and their outcomes by country, and adopt policies that distinguish extremist violence from political opposition designated as extremist.

For civil society and researchers

Civil society organisations are central to prevention and should be protected from the counter-terrorism regulation that constrains them. Researchers should test the propositions in African settings, where the evidence on algorithmic radicalisation is almost entirely absent: audit studies of recommendation systems in African languages, experimental studies of exposure effects, and studies of how movements such as ISWAP and al-Shabaab integrate online and offline mobilisation. Detection research (Berjawi et al., 2023) should prioritise multilingual and audio capability and should be evaluated against real-world outcomes rather than classification accuracy alone.

Limits

The paper is conceptual. It draws most of its algorithmic evidence from studies of Western platforms and audiences. It treats extremist movements as a category when jihadist, ethno-nationalist and other movements differ in their use of digital media. And it does not address the use of

© IIARJournal. All rights reserved.

artificial-intelligence tools by extremist movements themselves, which is an emerging concern.

5. CONCLUSION AND RECOMMENDATION

The rabbit-hole image of digital extremism is a poor guide for Africa. Recommendation algorithms shape exposure, but exposure is the least consequential of the three functions digital platforms perform for extremist movements; community and operation matter more, and operation has moved to channels moderation cannot reach. Content moderation in African languages and contexts is too weak to remove genuine extremist content and too strong in its removal of lawful speech, and the counter-extremism state that steps in where platforms fail can generate the grievance that sustains recruitment. The movements that kill most people on the continent mobilise through territory, governance and the failures of the state as much as through feeds.

This paper has offered a decomposition of platform functions, an account of moderation failure in African contexts, a warning about the counter-extremism state, and seven propositions that make the argument testable, anchored in Nigeria's long war with Boko Haram and ISWAP. The conclusion is that online counter-radicalisation has a place, but a smaller one than its prominence suggests. The decisive contest is offline: in the conduct of states, the quality of governance, and the credibility of the local voices who can answer extremist narratives in the languages in which they are told.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper contributes a decomposition of the functions digital platforms perform for extremist movements, exposure, community and operation, and shows that recommendation algorithms contribute most to the first, which matters least for violence, and least to the third, which has migrated to channels that content moderation cannot reach.

It further contributes an account of why content moderation fails in African contexts and a warning about the counter-extremism state, giving governments, platforms and civil society a basis for weighting counter-radicalisation effort toward the offline drivers of mobilisation, governance, grievance and state conduct, that the online frame obscures.

7. REFERENCES

Ayalew, Y. E. (2025a). Linguistic blind spots in platform content moderation in Ethiopia: The case of Facebook. Modern Languages Open, 2025(1), 11. https://doi.org/10.3828/mlo.v0i0.468

Ayalew, Y. E. (2025b). A third-world critique of the international human rights-based approach to content moderation. Transnational Legal Theory, 16(4), 546–577. https://doi.org/10.1080/20414005.2025.2523184

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

Berjawi, O., Fenza, G., & Loia, V. (2023). A comprehensive survey of detection and prevention approaches for online radicalization: Identifying gaps and future directions. IEEE Access, 11, 120463–120491. https://doi.org/10.1109/access.2023.3326995

Haroon, M., Wojcieszak, M., Chhabra, A., Liu, X., Mohapatra, P., & Shafiq, Z. (2023). Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations. Proceedings of the National Academy of Sciences, 120(50), e2213020120. https://doi.org/10.1073/pnas.2213020120

Jost, P., & Dogruel, L. (2023). Radical mobilization in times of crisis: Use and effects of appeals and populist communication features in Telegram channels. Social Media + Society, 9(3), 20563051231186372. https://doi.org/10.1177/20563051231186372

Khaemba, S. W. (2026). The role of social enterprise in preventing and countering violent extremism in Sub-Saharan Africa. Journal of Entrepreneurship and Public Policy, 15(1), 111–128. https://doi.org/10.1108/jepp-06-2024-0106

Lakomy, M. (2023). Why do online countering violent extremism strategies not work? The case of digital jihad. Terrorism and Political Violence, 35(6), 1261–1298. https://doi.org/10.1080/09546553.2022.2038575

Liu, N., Hu, X. E., Savas, Y., Baum, M. A., Berinsky, A. J., Chaney, A. J. B., Lucas, C., Mariman, R., de Benedictis-Kessner, J., Guess, A. M., Knox, D., & Stewart, B. M. (2025). Short-term exposure to filter-bubble recommendation systems has limited polarization effects: Naturalistic experiments on YouTube. Proceedings of the National Academy of Sciences, 122(8), e2318127122. https://doi.org/10.1073/pnas.2318127122

McMinimy, K., Winkler, C. K., Lokmanoglu, A. D., & Almahmoud, M. (2023). Censoring extremism: Influence of online restriction on official media products of ISIS. Terrorism and Political Violence,

© IIARJournal. All rights reserved.

35(4), 971–987. https://doi.org/10.1080/09546553.2021.1988938

Nicholas, G., & Bhatia, A. (2023). Toward better automated content moderation in low-resource languages. Journal of Online Trust and Safety, 2(1). https://doi.org/10.54501/jots.v2i1.150

Njoku, E. T. (2025). Counter-terrorism measures and civil society organisations: A review of literature and agenda for future research. Journal of Civil Society, 21(2), 205–235. https://doi.org/10.1080/17448689.2025.2451928

Ojewale, O. (2025). Triad of violence: Evolution and dynamics of conflict, terrorism, banditry and armed groups in northwest Nigeria. Dynamics of Asymmetric Conflict, 18(1), 18–43. https://doi.org/10.1080/17467586.2024.2356511

Okpaleke, F. N., Nwosu, B. U., Okoli, C. R., & Olumba, E. E. (2023). The case for drones in counter-insurgency operations in West African Sahel. African Security Review, 32(4), 351–367. https://doi.org/10.1080/10246029.2023.2217158

Osisanwo, A. (2025). Manipulative and recruitment strategies in Boko Haram terrorist statements. Journal of Asian and African Studies, 60(5), 3303–3318. https://doi.org/10.1177/00219096241228795

Rothut, S., Schulze, H., Hohner, J., Greipl, S., & Rieger, D. (2026). Radicalization and the internet: 25 years of (online) radicalization research. The Communication Review. Advance online publication. https://doi.org/10.1080/10714421.2026.2658935

Shin, D., & Jitkajornwanich, K. (2024). How algorithms promote self-radicalization: Audit of TikTok's algorithm using a reverse engineering method. Social Science Computer Review, 42(4), 1020–1040. https://doi.org/10.1177/08944393231225547

Stoddard, E. (2023). Competitive control? 'Hearts and minds' and the population control strategy of the Islamic State West Africa Province. African Security, 16(1), 32–60. https://doi.org/10.1080/19392206.2023.2192158

Weimann, G., & Pack, A. (2026). TamTam: The online drums of hate. Studies in Conflict & Terrorism, 49(2), 145–160. https://doi.org/10.1080/1057610x.2023.2225275

Whittaker, J., & Craanen, A. (2025). The unintended consequences of the removal of terrorist content and the case of Bitchute. Studies in Conflict & Terrorism. Advance online publication. https://doi.org/10.1080/1057610x.2025.2595843

© IIARJournal. All rights reserved.