

 **Global Journal of Arts Humanity and Social Sciences**
ISSN: 2583-2034
Abbreviated key title: Glob.J.Arts.Humanit.Soc.Sci
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link: <https://gsarpublishers.com/journal-gjahss-home/>



Volume - 6 | Issue - 10 | October 2026 | Total pages 1047-1057 | DOI: 10.5281/zenodo.23558690

ANALYSIS OF HATE SPEECH REGULATION ON TIKTOK SKITS IN NIGERIA

By

Atuluku, S¹, Haruna, S.A.¹, Wonah, P.O², Abbah, E.O³, Yusufu, G.E⁴ and Onuh-Psalm, O.K⁵

¹Department of Mass Communication, Nassarawa State University, Keffi, Nassarawa State, Nigeria

²Department of English and Language Arts Education, University of Calabar, Cross Rivers State, Nigeria

³Department of English and Literary Studies, University of Lagos Annex, Akoka, Lagos State, Nigeria

⁴Department of Philosophy, Prince Abubakar Audu University, Anyigba, Kogi State, Nigeria

⁵Department of Mass Communication, University of Nigeria, Nsukka, Enugu State, Nigeria



Article History

Received: 05- 10- 2026

Accepted: 13- 10- 2026

Published: 15- 10- 2026

Corresponding author
Atuluku, S

Abstract

This study examines how hate speech is regulated on TikTok skits in Nigeria's digital media space, asking whether the country's regulatory frameworks are adequate and how well TikTok's own content moderation copes with hate speech hidden inside short comedic videos. Two theories anchor the analysis: Agenda-Setting Theory (McCombs & Shaw, 1972; McCombs, 2014), which explains how unmoderated hate speech shapes what the public pays attention to, and Governance Theory (Rhodes, 1996; Flew, Martin, & Suzor, 2019), which accounts for the many actors involved in digital content governance. A systematic documentary research approach was adopted, drawing on academic journal articles, institutional reports, government policy documents, TikTok's transparency records, civil society monitoring reports, and international regulatory frameworks published between 2019 and 2024. Two things emerged. First, Nigeria's existing regulatory instruments, including the Cybercrimes (Prohibition, Prevention, etc.) Act 2015, the Nigerian Broadcasting Commission Code, and the NITDA Guidelines, are each inadequate on their own and, taken together, still fail to address TikTok-specific skit content. Second, TikTok's Community Guidelines show serious enforcement gaps: algorithmic moderation reportedly misses an estimated 88% of ethnically charged Nigerian skit content, largely because the platform's language-processing tools were never built for indigenous Nigerian languages. Five structural gaps drive this failure: legislative fragmentation, poor inter-institutional coordination, weak platform localisation, thin government-platform cooperation, and redress channels that ordinary users cannot reach. The study proposes a multi-layered governance framework to close these gaps and contributes to digital media governance scholarship by documenting regulatory failure in a major African digital market and outlining a co-regulatory model suited to linguistically diverse, low-resource-language contexts.

Keywords: hate speech, TikTok skits, content moderation, Nigeria, digital regulation, platform governance, documentary research

Introduction

Social media has reshaped global communication, allowing billions of people to create, share, and consume content at a speed and scale that would have been unimaginable a generation ago. One troubling consequence of this shift is the amplification of hate speech, expression that targets individuals or groups on the basis of

protected characteristics such as ethnicity, religion, gender, or national origin. Regulating hate speech online has become one of the most contested issues in communication law, human rights discourse, and media policy worldwide. Gagliardone, Gal, Alves, and Martinez (2020) point out that the lack of harmonised international standards for online hate speech has left significant



enforcement gaps, particularly across the Global South, where institutional capacity for digital governance tends to be limited. Underlying all of this is a basic tension between protecting freedom of expression and preventing speech that incites discrimination, hostility, or violence (Gagliardone et al., 2020).

Globally, platforms such as Facebook, YouTube, Twitter/X, and TikTok govern hate speech mainly through their own internal community standards, arrangements the literature often labels 'private governance' regimes, and ones that have attracted considerable scholarly scrutiny. Brown (2020) argues that self-regulation by tech companies is inherently limited by profit motives, inconsistent enforcement, and algorithmic bias that disadvantages non-English content. The United Nations Strategy and Plan of Action on Hate Speech (United Nations, 2020) goes further, arguing that platform-level responses alone are insufficient without state-level legislation binding platforms to specific obligations. Scholars and policymakers increasingly favour multi-stakeholder models that bring governments, civil society, platforms, and affected communities into a single governance process (Flew, Martin, & Suzor, 2019; Gorwa, Binns, & Katzenbach, 2020).

In Africa, the digital revolution has unfolded alongside deep social divisions along ethnic, religious, and political lines, creating fertile ground for online hate speech. Ethiopia, Kenya, South Africa, and Nigeria have each seen cases where social media content fed directly into inter-group tension and, at times, physical violence (Maina, 2021; Ekoh & George, 2021). The African Union's Declaration on Freedom of Expression and Access to Information in Africa (2019) calls on member states to build robust legal frameworks for online hate speech while still preserving democratic freedoms. Yet, as Olonade, Walters, Adebayo, and Ayeni (2020) point out, most African states have not translated these commitments into enforceable digital legislation, leaving a regulatory vacuum that platforms, whether deliberately or simply through inertia, exploit by under-enforcing their own community standards.

Nigeria offers a particularly compelling case. As Africa's most populous country and one of its largest digital markets, it had an estimated 35.6 million active social media users as of 2023 (Statista, 2023; DataReportal, 2024), and its digital public sphere is as volatile as it is vibrant. With over 250 recognised ethnic groups, dominated numerically by the Hausa-Fulani, Igbo, and Yoruba, and a long history of religious sectarianism and regional political tension, Nigeria is a place where online hate speech tends to map directly onto existing social fault lines, with real potential for offline harm. Adesanya, Chukwu, and Bello (2022) find that hate speech incidents on Nigerian social media spike during elections and moments of national crisis, with ethnic and religious targeting the most common categories. Despite this, Nigeria's legislative response has remained fragmented and inadequate, relying on general cybercrime statutes applied on an ad hoc basis rather than any dedicated hate speech law (Adewale & Okon, 2021; Ibrahim, 2023).

TikTok, the short-form video platform owned by ByteDance, has become one of the fastest-growing social platforms both globally and in Nigeria, where local creators are among the continent's most active. Its format, short, algorithmically curated skits usually set to music or trending audio, has made it especially popular with Nigerian youth, who use it for entertainment, social commentary, cultural expression, and, increasingly, political satire. That same format, however, has also become a vehicle for hate speech, particularly skits that mock, dehumanise, or stereotype ethnic, religious, or regional groups (Olanrewaju & Folarin, 2023). The comedic framing makes this a distinctive regulatory problem, since humour gives creators a kind of plausible deniability that makes it harder for both automated systems and trained human reviewers to flag such content as hate speech (Nwachukwu & Eze, 2024; Gorwa et al., 2020). TikTok's Community Guidelines formally prohibit content that promotes hatred against protected groups, but enforcement in the Nigerian context has been inconsistent, and culturally specific, linguistically nuanced hate speech routinely slips past both algorithmic and human detection.

It is against this backdrop of legislative inadequacy, institutional fragmentation, platform moderation failure, and heightened social vulnerability that this study analyses the regulatory frameworks governing hate speech on TikTok skits in Nigeria and assesses how effectively TikTok's content moderation actually manages hate speech in the Nigerian digital environment. In doing so, it addresses a gap in the existing scholarship and offers empirical and analytical groundwork for evidence-based regulatory reform.

Statement of the Problem

TikTok's rapid growth into a dominant social media platform in Nigeria has produced a significant, and so far underexamined, regulatory crisis. National and platform-level frameworks nominally exist to govern digital content, yet hate speech embedded in TikTok comedic skits keeps proliferating with little accountability or consequence. Three interconnected problems motivate this study.

First, Nigeria's existing regulatory frameworks are poorly suited to hate speech on platforms like TikTok. The main legislative instrument in use, the Cybercrimes (Prohibition, Prevention, etc.) Act 2015, was never designed with platform-specific content moderation in mind, and its provisions manage to be both too broad and not specific enough to guide hate speech governance on social media. Adewale and Okon (2021) show that without a dedicated hate speech statute, enforcement in Nigeria ends up reactive, politically contingent, and inconsistent. Regulatory bodies such as the Nigerian Broadcasting Commission (NBC), the National Information Technology Development Agency (NITDA), and the National Human Rights Commission (NHRC) have overlapping but poorly coordinated mandates, and platforms exploit the resulting institutional gaps by simply defaulting to their own community standards, standards that are themselves poorly calibrated to the Nigerian context.

Second, TikTok's content moderation infrastructure has proven ineffective at catching hate speech in Nigerian skit content.



Olanrewaju and Folarin (2023) found that roughly 78% of sampled Nigerian TikTok skits contained ethnically charged language, yet only 12% of that content was removed following algorithmic review, a shortfall traced to moderation tools that simply are not calibrated for Nigerian languages, dialects, or cultural registers. Because of this blind spot, hate speech delivered in Pidgin English, Yoruba, Igbo, Hausa, or other indigenous languages tends to slip through undetected (Nwachukwu & Eze, 2024; Access Now, 2023). TikTok's compliance with Nigerian government takedown requests, meanwhile, fell below 40% during the 2022 electoral period (Adesanya et al., 2022).

Third, Nigerian TikTok users affected by hate speech have almost no accessible avenue for redress. Access Now (2023) found that TikTok's appeals system was effectively out of reach for 83% of Nigerian users, largely because of language barriers and the absence of locally adapted support. Taken together, these three failures, legislative fragmentation, inadequate platform moderation, and thin user redress, add up to a compounded governance crisis that calls for careful scholarly analysis and evidence-based policy intervention.

Objectives of the Study

The specific objectives of this study are to:

1. Examine and evaluate the regulatory frameworks currently governing hate speech on TikTok skits in Nigeria, and determine how adequate they are individually and collectively.
2. Assess how effectively TikTok's content moderation mechanisms manage hate speech in Nigerian TikTok skits, drawing on empirical evidence from multiple primary and secondary sources.
3. Identify the structural gaps in current regulatory and platform-level governance, and propose an evidence-based, multi-layered regulatory framework for addressing hate speech on TikTok in Nigeria.

Conceptual Review

Hate Speech

Hate speech remains contested territory across communication studies, law, political science, and philosophy. Broadly, it refers to any expression, spoken, written, symbolic, visual, or audio-visual, that attacks, dehumanises, or incites hostility against a person or group on the basis of protected characteristics such as race, ethnicity, religion, gender, sexual orientation, or national origin (Brown, 2020; Walker, 2020). The breadth of this definition reflects how difficult it is to draw a clean line between protected expression and harmful speech, a tension that has shaped academic debate and policy across jurisdictions for decades. Walker (2020) frames this as an oscillation between content-based approaches, which focus on how injurious a message is, and effect-based approaches, which examine the actual harm caused to target groups or to social cohesion.

The United Nations (2020) defines hate speech as communication that attacks or uses pejorative or discriminatory language tied to

identity characteristics, noting that while international human rights law does not ban hate speech outright, it does require states to prohibit advocacy of national, racial, or religious hatred that amounts to incitement to discrimination, hostility, or violence. Distinguishing between speech that is merely offensive and speech that rises to actionable incitement sits at the centre of legal and regulatory frameworks everywhere. That distinction becomes far harder to draw online: the scale, speed, and algorithmic amplification of digital communication can turn even a localised expression of prejudice into viral content with far-reaching consequences (Flew et al., 2019; Gorwa et al., 2020).

In Nigeria, hate speech typically carries ethnic, religious, and regional undertones that mirror the country's social geography. Olonade, Walters, Adebayo, and Ayeni (2020) observe that Nigerian hate speech often works through coded language, cultural innuendo, and community-specific references that are legible to insiders but opaque to regulatory systems built around Western linguistic and cultural assumptions. Ibrahim (2023) documents how digital platforms amplify existing ethnic and religious cleavages, particularly around elections and periods of national security stress. Regulating hate speech in Nigeria, then, requires more than definitional clarity; it demands cultural competence and contextual sensitivity from regulatory authorities, platform governance actors, and moderation systems alike.

Content Moderation

Content moderation is the practice through which platforms, governments, or regulatory bodies review, filter, remove, or restrict user-generated content that violates community standards, legal norms, or platform policy. It spans a wide range of practices, from automated algorithmic filtering to human review teams, crowd-sourced flagging, and government-mandated takedown orders (Gillespie, 2020; Roberts, 2019). Gorwa, Binns, and Katzenbach (2020) propose a three-part typology: automated moderation using machine learning and natural language processing, reactive moderation that responds to user reports after the fact, and proactive moderation that filters content before it is published. Each approach has its own limitations, and all three tend to underperform in multilingual, multicultural settings.

How well moderation works depends on several things working together: how technically sophisticated and linguistically comprehensive the detection algorithms are, how well-trained and culturally competent human reviewers are, how clear and context-specific the community guidelines are, and how accessible and fair the appeals process is for content removed by mistake (Roberts, 2019; Gillespie, 2020). Research consistently shows that moderation systems built primarily for English-speaking, Western contexts underperform elsewhere. Roberts (2019) documents the human cost of this work, noting that platforms often rely on contract workers in the Global South to review flagged content, workers who are frequently under-trained to judge culturally nuanced hate speech from regions and languages they do not know well. This gap shows up clearly in TikTok's moderation infrastructure, which has been criticised repeatedly for focusing on

English-language content while neglecting African linguistic diversity (Access Now, 2023; Ekoh & George, 2021).

For Nigerian TikTok skits specifically, moderation faces a compound problem. The comedic format gives creators layers of plausible deniability that disguise harmful content, while Nigeria's linguistic diversity, Pidgin English, Yoruba, Igbo, Hausa, and dozens of minority languages, creates detection blind spots that automated systems simply are not built to handle. Olanrewaju and Folarin (2023) argue that effective moderation in the Nigerian digital context requires a fundamentally different approach, one that combines local cultural expertise, indigenous-language natural language processing, and community-based flagging within a single moderation architecture. Without that kind of structural change, platform moderation remains, as Gillespie (2020) puts it, performative at best and complicit at worst in allowing online hate speech to continue.

Digital Media Regulation

Digital media regulation covers the laws, policies, institutional arrangements, and normative frameworks through which state and non-state actors attempt to govern what gets produced, circulated, and consumed on digital platforms. It spans everything from statutory legislation and independent regulatory agencies to platform self-regulation, co-regulatory arrangements, and international governance instruments (Flew, Martin, & Suzor, 2019; Mansell, 2020). Digital media governance has become one of the more complex and fast-moving areas of public policy globally, as states search for regulatory approaches that are effective, proportionate, and still protective of fundamental rights.

A recurring tension in this space lies between the territorial reach of national regulators and the transnational nature of major platforms. TikTok operates across borders, so it is nominally subject to the regulatory authority of every country in which it operates, but in practice it resists comprehensive national oversight through a combination of technical architecture, legal challenges, and regulatory arbitrage (Brown, 2020; Mansell, 2020). Nigeria feels this acutely: agencies such as NITDA and the NBC have asserted jurisdiction over digital platforms but lack the enforcement capacity, technical expertise, and diplomatic leverage needed to actually compel compliance from global tech companies. Maina (2021) identifies this mismatch between national authority and platform power as especially consequential in places with elevated social tension, since it allows unchecked algorithmic amplification of content that can contribute to real-world inter-group conflict.

Flew, Martin, and Suzor (2019), Gorwa, Binns, and Katzenbach (2020), and Mansell (2020) all arrive at a similar conclusion: effective digital media regulation must move past the binary of pure state regulation versus platform self-regulation, toward collaborative, multi-stakeholder co-regulatory models. For Nigeria, that means comprehensive, specifically targeted digital hate speech legislation with mandatory platform compliance; institutional strengthening of regulatory bodies with genuine digital expertise and enforcement power; formal co-regulatory frameworks that

require platforms to meet minimum content governance standards calibrated to local linguistic and cultural context; and accessible redress mechanisms for users whose rights are violated through platform-mediated harm.

Theoretical Framework

Agenda-Setting Theory

This study draws in part on Agenda-Setting Theory, first developed by Maxwell McCombs and Donald Shaw (1972) and later extended through a substantial body of scholarship, culminating in McCombs's (2014) second-level agenda-setting model. The theory holds that mass media do not tell people what to think so much as they shape which issues people consider important; the prominence media give an issue directly shapes how salient it becomes in public consciousness. Scholars such as Gilardi, Gessler, Kubli, and Müller (2022) and Scheufele and Tewksbury (2007) have extended this to the digital landscape, arguing that algorithmic curation and viral content distribution function as agenda-setting mechanisms in their own right, arguably even more powerful than traditional mass media in this respect.

This theory is useful here because it helps explain the downstream social consequences of poorly regulated hate speech on TikTok. When TikTok's moderation systems fail to catch and remove hate speech embedded in algorithmically amplified skit content, that content reaches a mass audience and, in effect, sets a social agenda, one that gradually normalises ethnic prejudice, inter-group hostility, and religious intolerance. The theory therefore offers a way of understanding not just that hate speech exists on the platform, but how, left unchecked, it can shape social perceptions, harden divisions between groups, and contribute to a cultural climate more conducive to discrimination and conflict in Nigeria (Flew et al., 2019; McCombs, 2014). Anchoring the study in Agenda-Setting Theory pushes the analysis beyond simply describing hate speech incidents, toward an explanatory account linking regulatory failure to measurable social harm.

Governance Theory

The study's second theoretical anchor is Governance Theory, developed by Rhodes (1996) and later applied to digital media regulation by Flew, Martin, and Suzor (2019) and Mansell (2020). In its contemporary form, the theory treats governance not as something states do alone, but as a networked process involving multiple interdependent actors, governments, regulatory agencies, private companies, civil society organisations, and affected communities, who together shape the rules governing a given social domain. That framing fits digital platform governance well, since neither pure state regulation nor pure platform self-regulation has proven sufficient, and effective governance depends on structured collaboration across sectors.

Applied to the Nigerian TikTok context, Governance Theory helps diagnose the structural causes of the dual governance crisis this study identifies, and it offers a way of thinking through the multi-layered interventions needed to address it. It illuminates what happens when the relevant actors, the Nigerian government, NBC, NITDA, NHRC, TikTok, and civil society, fail to coordinate their



governance functions, and it provides normative grounding for the co-regulatory framework proposed here. Combining Governance Theory with Agenda-Setting Theory gives the study both explanatory depth, in accounting for why the regulatory crisis persists, and prescriptive capacity, in identifying the institutional and normative changes needed to resolve it.

Methodology

This study used a systematic documentary research methodology, a qualitative secondary data approach built around the structured collection, critical review, and thematic analysis of existing documents, records, and institutional data relevant to the research problem (Bowen, 2020; Prior, 2016). The documentary method suited the study's objectives well: examining regulatory frameworks and evaluating moderation effectiveness across multiple institutional sources called for critical review of existing material rather than new primary data collected through surveys or interviews.

The study drew on a deliberately varied set of documentary sources, chosen for breadth, rigour, and relevance: (1) peer-reviewed journal articles published between 2019 and 2024, from journals including the Journal of African Media Studies, African Communication Research, Journal of Digital Media & Policy, and the African Human Rights Law Journal; (2) primary legislative and policy documents, including the Cybercrimes (Prohibition, Prevention, etc.) Act (2015), the Nigerian Broadcasting Commission Code (2020), the NITDA Guidelines (2022), and the

Constitution of the Federal Republic of Nigeria (1999, as amended); (3) TikTok's Community Guidelines (2022–2024) and Transparency Reports (2023); (4) civil society monitoring reports, including Access Now (2023); (5) international regulatory frameworks, including the United Nations Strategy and Plan of Action on Hate Speech (2020) and the African Union Declaration on Freedom of Expression (2019); and (6) statistical data from Statista (2023) and DataReportal (2024) on Nigerian social media usage.

Data were analysed using thematic content analysis (Braun & Clarke, 2020). Recurring themes around regulatory framework adequacy, moderation effectiveness, institutional coordination, and platform localisation were identified, coded, categorised, and interpreted against the study's three objectives. Agenda-Setting Theory and Governance Theory guided the interpretive lens throughout, shaping both how the documentary data were read and the normative standards against which regulatory and platform performance were judged. To keep interpretation as rigorous as possible, the study triangulated across multiple types of documentary sources, so that findings on regulatory adequacy and moderation effectiveness rested on convergent evidence from legislative texts, empirical research, civil society monitoring data, and platform transparency records.

Results

Objective One: Existing Regulatory Frameworks Governing Hate Speech on TikTok Skits in Nigeria

Table 1: Regulatory Frameworks Governing Hate Speech on TikTok Skits in Nigeria

Regulatory Framework	Scope/Coverage	Key Provisions on Hate Speech	Applicability to TikTok Skits	Principal Limitations
Cybercrimes (Prohibition, Prevention, etc.) Act 2015	All digital communications and online platforms in Nigeria	Section 24: criminalises offensive, threatening, or hateful online communications; penalty up to 3 years imprisonment and/or fine	Directly applicable; TikTok video skits shared online fall within its digital communication jurisdiction	General provisions not tailored to social media skit content; enforcement selective and politically influenced; no hate speech-specific definitions (Adewale & Okon, 2021)
Nigerian Broadcasting Commission (NBC) Code (2020)	Licensed broadcast content on Nigerian platforms	Prohibits content inciting ethnic, religious, or racial hatred; requires prior content review for licensed broadcasters	Limited applicability; TikTok is a social media platform, not a licensed broadcaster, placing it outside NBC's primary mandate	Outdated analogue broadcasting framework; no social media jurisdiction; NBC lacks enforcement capacity over digital platforms (NBC, 2020)
National Information Technology Development Agency (NITDA) Guidelines (2022)	Digital services and data processing entities operating in Nigeria	Requires platforms to respond to government content takedown requests within 24 hours; mandates appointment of local compliance officers	Applicable to TikTok as a digital service provider; however, compliance by TikTok remains inconsistent	No specific hate speech provisions; enforcement against multinational platforms hampered by limited diplomatic leverage and technical capacity (NITDA, 2022)

Regulatory Framework	Scope/Coverage	Key Provisions on Hate Speech	Applicability to TikTok Skits	Principal Limitations
TikTok Community Guidelines (2022–2024)	All TikTok users globally	Prohibits dehumanising content targeting protected groups; employs AI-based automated moderation and human review escalation	Primary operational content governance mechanism; applies to all Nigerian TikTok users and creators	Algorithmic systems lack Nigerian language and cultural competence; enforcement inconsistent; appeals process inaccessible to most Nigerian users (Access Now, 2023)
Constitution of the Federal Republic of Nigeria, Section 39 (1999, as amended)	All Nigerian citizens	Guarantees freedom of expression; permits restrictions for public order, morality, and national security	Creates constitutional tension with hate speech regulation; frequently invoked by content creators to challenge content removals	Broad freedom of expression guarantee without digital-specific hate speech exceptions; invites legal challenges against regulatory action (CFRN, 1999)
National Human Rights Commission (NHRC) Act 2010 (as amended)	Human rights violations by public and private actors in Nigeria	Empowers NHRC to investigate and mediate complaints of rights violations, including hate speech-related harm	Tangentially applicable; NHRC can receive and investigate complaints from individuals harmed by TikTok hate speech content	Lacks enforcement powers against digital platforms; no statutory mandate for social media hate speech specifically (NHRC, 2010)

Source: Nigerian Broadcasting Commission (2020); Federal Republic of Nigeria (2015, 1999); TikTok (2023); NITDA (2022); NHRC (2010).

Table 1 paints a picture of a fragmented and structurally inadequate regulatory landscape. Six frameworks apply to varying degrees, spanning national legislation, regulatory agency mandates, and TikTok's own Community Guidelines, but three problems stand out. First, none of them was built with social media skit content in mind: the Cybercrimes Act (2015) covers digital communications generically, the NBC Code was designed for licensed broadcasters, NITDA's Guidelines address data governance and platform compliance rather than content, the NHRC Act deals with human rights complaints broadly, and TikTok's own Guidelines apply a global standard that does not account for Nigerian linguistic or

cultural specifics. Second, these frameworks sit in separate institutional silos with no coordinated inter-agency mandate, which creates jurisdictional ambiguity that TikTok exploits by simply defaulting to its own, poorly localised community standards. Third, the constitutionally guaranteed freedom of expression under Section 39, essential as a democratic safeguard, sits in constant tension with hate speech regulation, and without a specific hate speech statute, that tension provides recurring grounds for legal challenges against regulatory action.

Objective Two: Effectiveness of TikTok's Content Moderation in Managing Hate Speech in Nigerian Skits

Table 2: Effectiveness of TikTok's Content Moderation on Hate Speech in Nigerian TikTok Skits (2020–2024)

Study/Report (Year)	Country/Context	Hate Speech Incidents Identified	Moderation Action Taken	Outcome/Effectiveness Rating	Gap Identified
TikTok Transparency Report (2023)	Global (Africa sub-data)	113.5 million videos removed globally for community guideline violations; breakdown by country not disclosed for Nigeria	Automated AI flagging with human review escalation; Africa received <2% of total moderation resources	Low specificity for Nigerian context; African content receives disproportionately fewer moderation resources	Absence of Nigeria-specific enforcement data
Olanrewaju & Folarin (2023)	Nigeria (TikTok platform)	78% of sampled Nigerian TikTok skits (n=320) contained ethnically charged language or imagery	Only 12% of flagged ethnically charged content removed by platform algorithms within 72 hours	Very Low 88% moderation failure rate on ethnic hate speech content	Platform algorithmic blindspot for Nigerian languages and cultural codes

Study/Report (Year)	Country/Context	Hate Speech Incidents Identified	Moderation Action Taken	Outcome/Effectiveness Rating	Gap Identified
Nwachukwu & Eze (2024)	Nigerian universities (n=412 students)	62% of respondents encountered hate speech in TikTok comedy skits at least monthly; 34% encountered it weekly	User reporting activated by 71% of respondents; average platform response time: 72 hours; 58% of reports received no moderation action	Low majority of user reports generated no platform response within standard timeframe	Absence of mandatory national moderation response standards
Adesanya, Chukwu & Bello (2022)	Nigeria (2022 general election period)	Documented hate speech surges on TikTok and Twitter/X during electoral campaigns; ethnic targeting constituted 67% of incidents	Nigerian government issued 43 content takedown requests to TikTok; platform compliance rate: 38%	Very Low TikTok compliance with government takedown requests below 40%	Absence of legally binding platform cooperation obligations
Access Now (2023)	West Africa: Nigeria, Ghana, Senegal	TikTok ranked second (after Facebook) for hate speech prevalence in Nigeria; 3,200+ incidents documented across three-month monitoring period	Moderation inconsistent; appeals process inaccessible to 83% of Nigerian users due to language barriers and absence of local support infrastructure	Low appeals process inaccessible to most Nigerian users; removal inconsistency documented across monitoring period	No localized moderation infrastructure or Nigerian-language appeals mechanism
Gagliardone et al. (2020)	Sub-Saharan Africa (Nigeria, Kenya, Ethiopia, South Africa)	Ethnic and religious hate speech disproportionately prevalent on video-sharing platforms; platform moderation systematically underperforms in African linguistic contexts	Platform-level responses rated inadequate across all four countries; government regulatory frameworks insufficient to compel platform compliance	Low across all sampled countries structural deficits in African-context moderation identified as systemic	Absence of pan-African digital governance frameworks with enforcement capacity
Maina (2021)	Africa (pan-continental regulatory review)	Documented correlation between unmoderated social media hate speech and offline inter-group violence in Nigeria and Ethiopia	Government-level regulatory interventions found to be reactive and inconsistently applied; no evidence of proactive platform-government collaboration	Very Low regulatory responses characterised as consistently behind the pace of platform-mediated hate speech escalation	Absence of proactive state-platform regulatory cooperation frameworks

Source: TikTok (2023); Olanrewaju & Folarin (2023); Adesanya et al. (2022); Access Now (2023); Nwachukwu & Eze (2024); Gagliardone et al. (2020); Maina (2021).

Table 2 shows a consistent pattern of moderation failure across the studies and reports examined. Whichever measure is used, the share of flagged content actually removed (12%, Olanrewaju & Folarin, 2023), government takedown compliance (38%, Adesanya et al., 2022), how often users report encountering unmoderated

hate speech (62% monthly, Nwachukwu & Eze, 2024), how accessible the appeals process is (out of reach for 83% of Nigerian users, Access Now, 2023), or the broader structural picture from comparative African research (Gagliardone et al., 2020; Maina, 2021), the evidence points the same way: TikTok's moderation infrastructure is not up to the job in the Nigerian context. What ties

these findings together is the platform's failure at linguistic and cultural localisation, which keeps automated detection systems from recognising hate speech expressed through Nigerian languages, dialects, culturally specific references, and comedic codes. Taken together, this evidence suggests that TikTok's

moderation, as currently built and deployed, simply is not fit for purpose in Nigeria.

Objective Three: Structural Gaps and Proposed Regulatory Framework

Table 3: Structural Regulatory Gaps and Multi-Layered Framework for Hate Speech Governance on TikTok in Nigeria

Dimension	Current Status (Nigeria)	Identified Gap	Recommended Intervention
Legislative Framework	No dedicated digital hate speech statute; reliance on Cybercrimes Act 2015 (general provisions)	Absence of platform-specific hate speech definitions and enforcement obligations	Enact a comprehensive Digital Hate Speech and Content Governance Act with mandatory platform compliance provisions
Regulatory Institutional Coordination	Fragmented mandates across NBC, NITDA, NHRC, and NCC with no unified digital content oversight body	Jurisdictional overlap and institutional gaps exploited by platforms defaulting to generic community standards	Establish a National Digital Content Regulatory Authority with unified mandate and enforcement powers
Platform Localisation	TikTok moderation systems lack Nigerian language and cultural competence; no indigenous language AI tools deployed	88% moderation failure rate on ethnically charged Nigerian content (Olanrewaju & Folarin, 2023)	Mandate TikTok to develop and deploy Nigerian-language NLP moderation tools as a condition of continued market operation
Government-Platform Cooperation	TikTok compliance rate with Nigerian government takedown requests at 38% (Adesanya et al., 2022)	No legally binding cooperation framework governing platform response to government content removal orders	Negotiate and formalise binding Memoranda of Understanding between Nigerian government and TikTok with compliance reporting requirements
User Redress Mechanisms	TikTok appeals process inaccessible to 83% of Nigerian users; average response time 72 hours (Access Now, 2023; Nwachukwu & Eze, 2024)	No national digital hate speech complaints tribunal or accessible user redress mechanism	Establish a National Digital Hate Speech Complaints Tribunal with authority to impose financial penalties on non-compliant platforms
Community-Based Moderation	No community-based flagging or co-moderation scheme involving Nigerian civil society in TikTok content governance	Automated moderation fails to capture culturally and linguistically nuanced hate speech; community expertise not harnessed	Develop a co-regulatory framework incorporating civil society organisations, digital rights groups, and language communities in moderation oversight

Source: Synthesised from Adewale & Okon (2021); Olanrewaju & Folarin (2023); Access Now (2023); Nwachukwu & Eze (2024); Flew et al. (2019); Gorwa et al. (2020).

Table 3 brings together the structural gaps identified across the documentary sources and maps them onto specific governance interventions. Five dimensions of failure emerge: legislative fragmentation, poor institutional coordination, weak platform localisation, thin government-platform cooperation, and redress channels that users cannot really access. Across all five, the evidence points to the same conclusion: this is not a case of one institution failing, but a compounded breakdown across multiple layers of governance at once. The interventions proposed here work across all five dimensions simultaneously, in line with the

multi-layered co-regulatory model that Flew, Martin, and Suzor (2019) and Gorwa, Binns, and Katzenbach (2020) put forward as the most theoretically sound and empirically supported approach to platform hate speech governance where regulatory fragmentation and limited institutional capacity are both at play.

Discussion

These findings confirm and build on existing scholarship on regulatory inadequacy in Nigeria's digital governance. The documentary evidence shows that the frameworks nominally governing hate speech on TikTok skits in Nigeria are fragmented,

poorly suited to the task, uncoordinated across institutions, and generally ill-equipped for social media content governance. This lines up with, and adds empirical weight to, the assessments of Adewale and Okon (2021) and Ibrahim (2023), who identify the absence of targeted digital hate speech legislation as the core structural deficit in Nigeria's digital governance architecture. This study extends that analysis by showing, through a multi-source documentary review, that the legislative gap is compounded by overlapping and poorly coordinated mandates across NBC, NITDA, NHRC, and the NCC, generating jurisdictional ambiguity that TikTok exploits by falling back on generic global community standards demonstrably ill-suited to Nigerian linguistic and cultural realities.

The findings also show that TikTok's moderation is not merely occasionally ineffective; it is structurally and systematically inadequate for the Nigerian context. The 88% moderation failure rate documented by Olanrewaju and Folarin (2023), the sub-40% government compliance rate found by Adesanya, Chukwu, and Bello (2022), and the 58% non-response rate on user reports documented by Nwachukwu and Eze (2024) together paint a picture of moderation systems that, in practice, barely function for Nigerian hate speech governance. This aligns with, and contextualises, Gorwa, Binns, and Katzenbach's (2020) theoretical and empirical work showing that automated moderation is structurally biased toward high-resource languages and Western cultural contexts, and underperforms in linguistically diverse, low-resource environments like Nigeria. The particular vulnerability of TikTok's skit format, where comedic framing gives creators plausible deniability that further undermines algorithmic detection, is a specific and significant dimension of moderation failure that existing literature has not fully addressed.

Read through the combined lens of Agenda-Setting Theory and Governance Theory, these findings carry serious, compounded social implications. From an agenda-setting perspective (McCombs & Shaw, 1972; McCombs, 2014), TikTok's algorithmic amplification of unmoderated hate speech in popular Nigerian skit content effectively sets a social agenda that gradually normalises ethnic prejudice, inter-group hostility, and religious intolerance in Nigeria's digital public sphere. Given the scale and speed of this agenda-setting effect, reaching millions of Nigerian users, including politically formative youth populations, the regulatory failure documented here is not merely an institutional problem; it is a matter of real social consequence. From a governance theory perspective (Rhodes, 1996; Flew et al., 2019), the findings show that the current arrangement, fragmented state regulation, inadequate platform self-governance, and no structured civil society participation, simply cannot provide adequate oversight of hate speech on TikTok in Nigeria. That governance vacuum is not incidental; it is a structural outcome of failing to build the multi-stakeholder co-regulatory architecture that contemporary digital governance theory identifies as necessary for effective platform oversight.

The study's third objective, proposing a multi-layered regulatory framework, builds directly on these theoretical and empirical

findings. The five-dimensional framework in Table 3 reflects insights from both theories: from Agenda-Setting Theory, the urgency of addressing moderation failures that allow harmful content to reach mass algorithmic amplification; and from Governance Theory, the need for an institutional architecture that distributes governance responsibility across state, platform, and civil society actors in a structured, accountable, and locally grounded way. The framework's specific proposals, a comprehensive Digital Hate Speech and Content Governance Act, a unified National Digital Content Regulatory Authority, mandatory TikTok localisation obligations, binding memoranda of understanding for government-platform cooperation, a National Digital Hate Speech Complaints Tribunal, and a co-regulatory framework that brings in civil society, together address the systemic governance failures documented here, and are broadly consistent with best-practice models emerging from South Africa, Kenya, and the European Union's Digital Services Act (Ekoh & George, 2021; European Commission, 2022).

Conclusion

This study has carried out a systematic, theoretically grounded analysis of the regulatory frameworks governing hate speech on TikTok skits in Nigeria, alongside a critical assessment of how effectively TikTok's content moderation actually manages such content. Drawing on a multi-source documentary corpus spanning legislation, empirical research, civil society monitoring, and platform transparency data, and anchored jointly in Agenda-Setting Theory and Governance Theory, the study arrives at three main findings.

First, Nigeria's national regulatory frameworks are individually and collectively inadequate for social media hate speech governance, marked by legislative fragmentation, institutional coordination gaps, and the absence of platform-specific obligations. Second, TikTok's content moderation systems are structurally and systematically ill-equipped to detect, remove, or deter hate speech in Nigerian skit content, with documented failure rates reaching 88% on ethnically charged content and government-request compliance below 40%. Third, these two failures are not independent institutional problems but reflect a compounded, systemic governance crisis, one that produces algorithmic amplification of harmful content with measurable potential for social harm, and that can only be addressed through a comprehensive, multi-layered, co-regulatory framework operating simultaneously across legislative, institutional, platform, cooperation, and redress dimensions.

Recommendations

Based on the findings and analysis of this study, the following recommendations are advanced:

1. The Nigerian government should enact a comprehensive Digital Hate Speech and Content Governance Act that directly addresses social media platforms, establishing precise legal definitions of hate speech calibrated to Nigeria's sociolinguistic context, binding enforcement mechanisms, mandatory platform compliance

obligations, and proportionate penalties for non-compliance.

2. The Nigerian government should establish a unified National Digital Content Regulatory Authority with a consolidated mandate, dedicated technical expertise, and enforcement powers over digital platforms, replacing the current fragmented arrangement across NBC, NITDA, and NHRC, which creates jurisdictional ambiguity that platforms exploit.
3. TikTok should be required, as a condition of continued market access in Nigeria, to develop and deploy culturally localised content moderation tools with natural language processing capability for Nigerian indigenous languages including Yoruba, Igbo, Hausa, and Pidgin English and to report quarterly on moderation performance metrics specific to Nigerian content.

References

1. Access Now. (2023). Hate speech in West African digital spaces: A monitoring report. Access Now Digital Rights Organization. <https://www.accessnow.org/west-africa-hate-speech-report>
2. Adesanya, A., Chukwu, M., & Bello, R. (2022). Social media hate speech and electoral violence in Nigeria. *Journal of African Media Studies*, 14(2), 189–207. https://doi.org/10.1386/jams_00085_1
3. Adewale, S., & Okon, P. (2021). Legislative gaps in Nigeria's digital hate speech governance: A critical review. *Nigerian Journal of Communication*, 18(1), 45–68.
4. African Union. (2019). Declaration on freedom of expression and access to information in Africa. African Commission on Human and Peoples' Rights.
5. Bowen, G. A. (2020). Documentary research as a qualitative strategy. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
6. Braun, V., & Clarke, V. (2020). *Thematic analysis: A practical guide*. SAGE Publications.
7. Brown, A. (2020). *Hate speech law: A philosophical examination*. Routledge.
8. Constitution of the Federal Republic of Nigeria. (1999, as amended). Federal Republic of Nigeria Official Gazette.
9. DataReportal. (2024). Digital 2024: Nigeria. DataReportal Global Digital Insights. <https://datareportal.com/reports/digital-2024-nigeria>
10. Ekoh, P. C., & George, E. O. (2021). Comparative social media regulation and hate speech management in Africa: A five-country analysis. *African Human Rights Law Journal*, 21(2), 445–469. <https://doi.org/10.17159/1996-2096/2021/v21n2a18>
11. European Commission. (2022). Digital Services Act: Regulation (EU) 2022/2065 of the European Parliament and of the Council. Official Journal of the European Union.
12. Federal Republic of Nigeria. (2015). Cybercrimes (Prohibition, Prevention, etc.) Act 2015. Federal Republic of Nigeria Official Gazette.
13. Flew, T., Martin, F., & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, 10(1), 33–50. https://doi.org/10.1386/jdmp.10.1.33_1
14. Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2020). Countering online hate speech. UNESCO Publishing.
15. Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2022). Social media and political agenda setting. *Political Communication*, 39(1), 39–60. <https://doi.org/10.1080/10584609.2021.1910390>
16. Gillespie, T. (2020). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
17. Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
18. Ibrahim, A. (2023). Digital platforms, hate speech, and electoral violence in Nigeria: The 2023 general election. *African Communication Research*, 12(2), 101–128.
19. Maina, G. (2021). Hate speech online and offline violence in Africa: Regulatory responses and their limitations. *African Human Rights Law Journal*, 21(1), 113–139. <https://doi.org/10.17159/1996-2096/2021/v21n1a5>
20. Mansell, R. (2020). Digital intermediaries, regulatory failure, and the turn toward co-regulation. *Information, Communication & Society*, 23(6), 823–839. <https://doi.org/10.1080/1369118X.2020.1718173>
21. McCombs, M. (2014). *Setting the agenda: Mass media and public opinion* (2nd ed.). Polity Press.
22. McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187. <https://doi.org/10.1086/267990>
23. National Human Rights Commission Act. (2010, as amended). Federal Republic of Nigeria Official Gazette.
24. National Information Technology Development Agency. (2022). NITDA guidelines for social media platforms operating in Nigeria. Federal Republic of Nigeria.
25. Nigerian Broadcasting Commission. (2020). *Nigerian Broadcasting Commission code* (Revised ed.). Nigerian Broadcasting Commission.
26. Nwachukwu, C., & Eze, F. (2024). Youth exposure to hate speech on TikTok: Evidence from Nigerian universities. *African Communication Research*, 11(1), 77–99.
27. Olanrewaju, T., & Folarin, B. (2023). Ethnic hate speech and TikTok content moderation in Nigeria: A content



- analysis. *Journal of Communication and Media Research*, 15(2), 201–218.
28. Olonade, O., Walters, G., Adebayo, A., & Ayeni, C. (2020). Hate speech, social media, and ethnic conflict in Nigeria. *Covenant Journal of Communication*, 7(1), 1–19.
 29. Prior, L. (2016). *Using documents in social research* (2nd ed.). SAGE Publications.
 30. Rhodes, R. A. W. (1996). The new governance: Governing without government. *Political Studies*, 44(4), 652–667. <https://doi.org/10.1111/j.1467-9248.1996.tb01747.x>
 31. Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.
 32. Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1), 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>
 33. Statista. (2023). Number of social media users in Nigeria 2023. Statista Research Department. <https://www.statista.com/statistics/1176058/social-media-users-nigeria/>
 34. TikTok. (2023a). TikTok community guidelines (2022–2024). ByteDance. <https://www.tiktok.com/community-guidelines>
 35. TikTok. (2023b). TikTok transparency report 2023. ByteDance. <https://www.tiktok.com/transparency/en-us/2023/>
 36. United Nations. (2020). *United Nations strategy and plan of action on hate speech*. United Nations.
 37. Walker, S. (2020). *Hate speech and freedom of speech in America*. Cambridge University Press.