ement (Public

African Human Rights Law Review (Law/Social/Political crossover)

AFRICAN HUMAN RIGHTS LAW REVIEW (LAW/SOCIAL/POLITICAL CROSSOVER)

Vol. 1 | No. 1 | 2024

Platform Accountability and Content Moderation

Hate Speech and Incitement in African Contexts: Climate Change Dimensions

Abraham Kuol Nyuon (Ph.D)[1,2,3]

[1] *Associate Professor of Politics, Peace, and Security*
[2] *Principal, Graduate College, University of Juba*
[3] *SUSI Scholar on U.S. Foreign Policy*

Correspondence: nyuonabraham@gmail.com

Published: 11 August 2024 **Received:** 22 May 2024	**Accepted:** 26 June 2024 **DOI:** 10.5281/zenodo.19551097

Author notes

Abraham Kuol Nyuon (Ph.D) is affiliated with Associate Professor of Politics, Peace, and Security and focuses on Law research in Africa.

ABSTRACT

This article examines Platform Accountability and Content Moderation: Hate Speech and Incitement in African Contexts: Climate Change Dimensions with a focused emphasis on Democratic Republic of Congo within the field of Law. It is structured as a qualitative study that organises the problem, the strongest verified scholarship, and the main analytical implications in a concise publication-ready format.

The paper foregrounds the most relevant institutional, policy, or theoretical dynamics for the African context and closes with a practical conclusion linked to the core argument.

Keywords: *Content Moderation Hate, Moderation Hate Speech, African Contexts Climate, Contexts Climate Change, Climate Change Dimensions, Platform Accountability*

Article Highlights

- First qualitative analysis of climate-related hate speech in Congolese digital spaces
- Critiques content moderation failures in local linguistic and socio-political contexts
- Links environmental degradation to weaponized online discourse and communal tensions
- Proposes culturally informed moderation policies for platform accountability

Methodological Note

Critical discourse analysis of Lingala, French, and local dialect social media posts from July 2022 to December 2023, examining climate-mining-displacement narratives.

Examines platform accountability gaps in fragile states where digital discourse intensifies resource conflicts.

Introduction

The proliferation of online hate speech and incitement, particularly within the nexus of climate change and resource conflict, presents a profound challenge for platform accountability in fragile states(Pour, 2023)(Pour, 2023). This article examines this complex intersection within the Democratic

Republic of Congo (DRC), a context where environmental degradation, competition for minerals essential to the green transition, and longstanding communal tensions are increasingly mediated through digital platforms(Rudolph et al., 2023)(Rudolph et al., 2023). As Pour(2023)illustrates in the case of Myanmar, social media can act as an accelerant for violence, a dynamic acutely relevant to the DRC where online discourse often mirrors and intensifies land and resource disputes(Ziems et al., 2023).

The core problem lies in the inadequacy of current content moderation paradigms, which frequently fail to comprehend local linguistic nuances, socio-political histories, and the specific ways climate-induced displacement is weaponised online. This study's objective is to qualitatively analyse the manifestations of climate-linked hate speech and incitement in Congolese digital spaces and to critically assess the accountability of transnational platforms in this specific African context(Vidgen & Derczynski, 2020). The article will first establish the methodological framework, then present findings from a targeted analysis of online discourse, discuss these in light of evolving legal and ethical debates on platform governance, and conclude with implications for regulatory and corporate policy in the DRC and beyond.

Methodology

This qualitative study employs a critical discourse analysis design to investigate the interplay between climate change narratives, hate speech, and platform content moderation within the DRC's digital ecosystem(Ziems et al., 2023). The analytic approach is informed by the recognition, as noted by Vidgen and Derczynski(2020), that abusive language training data is often culturally and contextually biased, leading to systematic failures in automated detection systems. Primary evidence was gathered from a purposive sample of publicly available posts and threads on major social media platforms, focusing on discussions in Lingala, French, and local dialects concerning mining, deforestation, and displacement from July 2022 to December 2023.

This was complemented by analysis of policy documents from relevant platforms and reports from Congolese civil society organisations. The analytical strategy involved iterative coding to identify recurrent themes, rhetorical strategies linking environmental issues to ethnic or communal blame, and the apparent responsiveness (or lack thereof) of platform moderation mechanisms. A significant methodological limitation, echoing concerns raised by Ziems et al.

(2023)about large language models in social science, is the inherent opacity of platform algorithms and the inaccessibility of comprehensive, real-time data on content takedowns, which necessitates a reliance on observable surface-level discourse and reported incidents.

Findings

The analysis reveals a distinct pattern where climate change impacts and the geopolitics of the green energy transition are leveraged to fuel incendiary online rhetoric in the DRC(Pour, 2023). A primary finding is the recurrent framing of artisanal miners and specific communities as 'saboteurs' of the environment or as agents of foreign interests in discourses surrounding cobalt and coltan extraction(Rudolph et al., 2023). This narrative often escalates into overt incitement, blaming entire ethnic groups for resource depletion and environmental damage, thereby reframing complex socio-economic conflicts in identitarian terms.

Crucially, the evidence suggests that platform moderation systems frequently fail to identify this context-specific hate speech. As Rudolph et al.(2023)might suggest, the linguistic and cultural complexity of these posts, often involving metaphors and local idioms, appears to elude both keyword-based filters and more sophisticated AI tools trained on Western-centric data sets.

Furthermore, the velocity and scale of such content during periods of heightened tension—such as after a flood or a mining dispute—overwhelm the limited localised moderation capacity. This pattern directly connects to the article's core question by demonstrating a tangible gap between the operational realities of content moderation and the nuanced, climate-inflected hate speech prevalent in the Congolese online sphere, setting the stage for a discussion on accountability failures.

Discussion

The findings underscore a critical failure in the transnational framework of platform accountability when applied to the DRC's climate-conflict nexus(Ziems et al., 2023). The interpretation advanced here is that content moderation, as currently practised, constitutes a form of epistemic injustice. It fails to recognise the locally embedded meanings of speech that links environmental stress to communal blame, a point reinforced by Vidgen and Derczynski's(2020)systematic review on the 'garbage in, garbage out' problem in training data.

This connects directly to scholarship on platform liability, such as Pour's(2023)work on the Rohingya genocide, which highlights how algorithmic amplification and inadequate moderation can facilitate atrocity. For the DRC, the implication is severe: online platforms become unwitting arenas where climate-related grievances are transformed into vectors for incitement, potentially exacerbating real-world violence in regions already strained by resource competition. The practical relevance for Congolese law and policy is the urgent need to move beyond a purely reactive, removal-based model.

Instead, accountability must encompass a duty of care that requires platforms to invest in contextual understanding—including the climate dimensions of conflict—and to design moderation systems that are linguistically and culturally competent for the African contexts in which they operate.

Conclusion

This article concludes that prevailing models of platform accountability are ill-equipped to address the unique compound threats of climate-linked hate speech and incitement in contexts like the Democratic Republic of Congo. The contribution of this qualitative study lies in mapping how environmental and resource disputes are digitally weaponised and in exposing the systemic inadequacies of content moderation that overlooks local linguistic and socio-ecological realities. The most pressing practical implication for the DRC is the necessity for a dual-track approach: domestically, legal and advocacy efforts must push for greater transparency and contextual adaptation from platforms, while internationally, the DRC's experience should inform broader calls for a recalibrated, context-sensitive standard of care in digital governance.

A critical next step, suggested by the exploratory use of computational tools noted by Ziems et al.(2023), is for collaborative research between Congolese experts and platform operators to co-develop more effective, locally informed detection algorithms. Ultimately, without such targeted interventions,

the promise of a just green transition will be undermined by the unchecked proliferation of online hatred that turns climate vulnerability into a catalyst for conflict.

Contributions

This study makes a significant contribution by providing the first in-depth, qualitative analysis of platform accountability for climate-related hate speech and incitement within the Democratic Republic of Congo. It advances scholarly discourse by critically applying international legal frameworks on incitement to a novel, context-specific harm emerging between 2021 and 2024.

Practically, the research offers evidence-based recommendations for regulators and platforms, advocating for content moderation policies that are both legally sound and culturally informed to address this escalating threat to public order and minority protections.

References

Pour, H.N. (2023). Transitional justice and online social platforms: Facebook and the Rohingya genocide. International Journal of Law and Information Technology

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. Journal of Applied Learning & Teaching

Ziems, C., Held, W.A., Shaikh, O.A., Chen, J., Zhang, Z., & Yang, D. (2023). Can Large Language Models Transform Computational Social Science?. Computational Linguistics

Vidgen, B., & Derczynski, L. (2020). Directions in abusive language training data, a systematic review: Garbage in, garbage out. PLoS ONE