MATTHIAS C. KETTEMANN, ANRIETTE ESTERHUYSEN AND JOSEFA FRANCKE (EDS.)

Platform://Democracy

Research Report Africa

Platform://Democracy

Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Africa

edited by Matthias C. Kettemann, Anriette Esterhuysen and Josefa Francke,
LEIBNIZ INSTITUTE FOR MEDIA RESEARCH | HANS-BREDOW-INSTITUT, HAMBURG, GERMANY
HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY, BERLIN, GERMANY

Cite as: Kettemann, Matthias C.; Esterhuysen, Anriette; Francke, Josefa (eds.)(2023): Platform://Democracy – Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Africa. Hamburg: Verlag Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.86525

This publication is part of the project *Platform://Democracy: Platform Councils as Tools to Democratize Hybrid Online Orders*. The project was carried out by the Leibniz Institute for Media | Hans-Bredow-Institut, Hamburg, the Alexander von Humboldt Institute for Internet and Society, Berlin, and the Department of Theory and Future of Law of the University of Innsbruck und funded by Stiftung Mercator.

Publisher: Leibniz Institut für Medienforschung | Hans-Bredow-Institut (HBI)
Rothenbaumchaussee 36, 20148 Hamburg
Tel. (+49 40) 45 02 17-0, info@leibniz-hbi.de, www.leibniz-hbi.de

Contributors

Name(s)	Affiliation
Anriette Esterhuysen	Association for Progressive Communications
Khadijah El-Usman	Paradigm Initiative
Nashilongo Gervasius	Namibia University of Science and Technology, London School of Economis
Grace Githaiga	Kenya ICT Action Network
Kuda Hove	Independent researcher
Tomiwa Ilori	Centre for Humna Rights, University of Pretoria
Ephraim Percy Kenyanito	Queen Mary University of London, University of East London
Thobekile Matimbe	Paradigm Initiative
Berhan Taye	Independent researcher
Emmanuel Vitus	AW Free Foundation

Table of Contents

Introduction to the Research Clinic Africa

Anriette Esterhuysen
ASSOCIATION FOR PROGRESSIVE COMMUNICATION, JOHANNESBURG, SOUTH AFRICA

Social media has transformed the character, scope, and scale of public discourse and political participation in Africa. In many respects the promise of digitalisation for social and economic development in the region has not been realised, and the average overall internet penetration is still well below 50%. Differences in internet penetration between regions and countries are dramatic, for example, in January 2022 it ranged from 7.1% in the Central African Republic to 84% in Morocco.[1] Insufficient availability of infrastructure and the high cost of data - Most African internet users connect via mobile phones - and devices continue to limit people's ability to access the internet. But in spite of these barriers, Africans have taken to social media with great enthusiasm. Patterns of use vary from other parts of the world in some respects, for example, statistics for 2022 indicates that average WhatsApp use in Africa is 35% higher than the global average.[2] Facebook is the most widely used platform, followed by YouTube, Twitter and Instagram.[3]

More Africans use social media to communicate than any other form of communication. Phone calls are expensive, and so is Short Message Service (SMS). Short calls or messages sent over social media platforms cost less, even though the cost of data in Africa is still comparatively high. A recently published commercial media monitoring survey found that, from among more than 170 participating countries, "African countries rely more on social media as a marketing tool (74%) compared to the overall results of other regions (52%)."[4] An Afrobarometer survey released in early 2022 indicates that use of digital sources for news had nearly doubled since 2015 with "more than four in ten adults across 34 surveyed countries reporting that they turn to the Internet or social media at least a few times a week for news".[5] Reliance on social media for news is reinforced by widespread lack of confidence in print and broadcast media.[6] Twitter is used widely for political conversations. African leaders tend to be ambivalent, on the one hand using social media themselves, and on the other hand feeling threatened and readily resorting to shutdowns during elections or protests.

Social media has become part of the fabric of daily life in Africa, yet African internet users have had very little influence on how it is governed. At national level many governments have introduced poorly thought-through regulations under the auspices of targeting harmful content and mis- and disinformation. However, these laws have generally been more effective at stifling online freedom of expression than in achieving the stated objective of curbing false information.[7] Nor have Africans had

[1] https://www.statista.com/statistics/1124283/internet-penetration-in-africa-by-country/

[2] State of Social Media for Africa 2023, Meltwater, January2023. https://meltwater.cdn.prismic.io/meltwater/91f084cd-7cb6-45a2-b072-4e319783e585_StateofSocial_Africa.pdf

[3] https://www.itnewsafrica.com/2022/07/top-5-most-popular-social-media-platforms-in-africa/

[4] https://www.bizcommunity.com/Article/196/669/235086.html

[5] https://www.afrobarometer.org/publication/ad509-promise-and-peril-in-changing-media-landscape-africans-are-concerned-about-social-media-but-opposed-to-restricting-access/

[6] In their 2022 World Press Freedom Index, Reporters Without Borders rated press freedom in only seven African countries as being satisfactory. https://rsf.org/en/index

[7] Misinformation Policy in Sub-Saharan Africa: From Laws and Regulations to Media Literacy, Universeity of Westminster Press, 2021. https://www.uwestminsterpress.co.uk/site/books/m/10.16997/book53/

much influence on how global social media platforms are governed, as pointed out in most of the contributions below.

But this lack of influence is not one dimensional. The perspectives on platform power, public values, and the potential of social media councils and other forms of platform governance discussed in the eight short papers below provide examples of how Africans can have more agency in the governance of social media platforms.

In *Rethinking Platform democracy in Africa: A Nigerian case study*, Khadijah El-Usman argues that social media platforms have acquired legitimacy in Nigeria but as they are based outside Africa and African users have no say in how they are governed, the nuances of indigenous languages, regional and political contexts, and larger interests of African users are often missed. On the other hand, where governments seek to enforce their own platform governance rules they often conflict with national and international human rights principles. Therefore, there is a need to introduce meaningful participation by African users in Social Media Councils, and combine global principles with national and regional regulation that is attuned to local contexts. She explores how existing regional regulation in West Africa can be used to resolve platform related complaints.

Nashilongo Gervasius Nakale examines decisions of the Meta Oversight Board's decisions in the African context and concludes that it applies universal laws to African realities, without making a serious effort to apply applicable national or regional regulatory frameworks. Consultation with local stakeholders, when there is any, is limited. Global rights frameworks and values are privileged at the expense of local nuances and contexts. An alternative approach would be to pay more attention to national legislative and human rights frameworks, and regional African human rights instruments, and convene more regional consultations. This would complement the existing oversight process. By involving local and regional bodies including civil society organisations, regulatory bodies, and human rights oversight mechanisms at the African Union level the Oversight Board's decisions would become more respectful of national democratic laws and processes.

Kuda Hove points out the same problem in his paper *A Proposal on Voluntary Social Media Councils*. To overcome that, he proposes that local regulatory models used for the print media can be modified and applied to social media platforms. He proposes the idea of voluntary national level social media councils which could consist of civil society actors, academics, users and representatives of special interest groups and minorities, to close the gap between the global and the local, and provide a legitimately independent, participatory, and transparent way to ensure user participation in content moderation and the regulation of social media platforms in each country.

Emmanuel Vitus' *Platform Democracy, one size does not fit all: the case of GhanaWeb* also focuses on the tension between the global standards used by social media platforms and the need to defer to local contexts when moderating content. He presents GhanaWeb an example of balancing freedom of speech for users and global trends to moderate online content perceived as potentially harmful. It engages in several phases of technical and human review in order to identify, assess, and take action against content that potentially violates its Community Standards through using local languages. The case of GhanaWeb shows how local platforms in Africa are struggling, and in this case succeeding, to build a transparent and strong moderation system to enhance user expression/experience and also keep the platform safe and democratic. There is a need to balance moderation with censorship, publish content guidelines and policies, disclose moderation practices and appeals process, and disclose algorithms.

In her paper on *Elections and social media platforms in Kenya* Grace Githaiga argues that social media platforms failed to uphold their commitment to maintain election integrity and prevent the proliferation

of harmful content during Kenya's 2022 elections. Civil society groups felt that moderators did not understand regional languages and did not take into consideration cultural and societal settings. There is a need for mechanisms to encourage fact-checking, combat hate speech, and engage stakeholders, including citizens and users, to report offenses, and to participate actively. Civil society must be strategic in engaging media platforms and point out local challenges. There is a need for international partners to make resources available resources to Global South Civil Society organisations in an effort to support their work in campaigns of engagement.

Ephraim Percy Kenyanito's article, *Unpacking the Tensions: A Comparative Analysis of DNS Abuse Mitigation and its Impact on Business and Human Rights* looks at DNS abuse in different countries, which poses a significant threat to human rights, particularly the right to freedom of expression. While DNS abuse mitigation measures are necessary, they must be consistent with international human rights law and not result in censorship or violate other human rights. Human rights considerations need to be integrated into DNS abuse mitigation measures, together with promotion of transparency and accountability in the implementation of these measures, and regular assessments of registry policies. By taking these steps, DNS abuse cane be mitigated while protecting human rights.

In *Centering Victims is Imperative for Effective Remediation in Platform Governance* Thobekile Matimbe calls on social media platforms to adopt a victim-centred approach as key in interventions that promote human rights. They should be transparent about how they address human rights and the remedial steps they take to address victims of harmful online content. Consulting with victim support groups and clinical psychologists is part of engaging different stakeholders in addressing victim needs. Platforms need to cooperate with other accountability mechanisms within the different contexts and jurisdictions where harmful content originates.

Tomiwa Ilori tackles what is in effect the overall challenge of platform governance in Africa in his article *Building and strengthening rights-based social media platform governance in Africa through national human rights institutions.* He points out that the laws, policies and processes made by African governments to regulate online harms are often at variance with international human rights standards while social media platforms are distant from the contextual realities required to regulate online harms in African countries. National Human rights institutions (NHRIs) can provide a link here as they have a key responsibility to ensure that key international human rights instruments are considered and implemented in local contexts. However, African NHRIs face a number of challenges that could make playing these roles difficult. They need to become involved in ensuring a rights-based approach to social media platform governance, build and strengthen the capacity of duty-bearers to meet their obligations while also encouraging rights holders to claim their rights. Some of the ways through which such a role can be played include strategic collaboration, policy advocacy, capacity building and radical participation.

Four themes cut through all the papers. First, the tension between using global standards as a basis for decision-making on whether content should stay or go, versus taking local contexts, cultures, values, and regulatory environments into account. Second, and linked to the first tension, is the difficulty of finding a balance between recognition and respect of local authorities and values without endorsing censorship, bias, or undue restrictions on human rights. Third is the challenge to achieve more and deeper, engagement from the people and institutions in Africa on how global governance frameworks are designed and applied. This involves creating more awareness of social media councils in the region, and investing in increasing African participation in their processes. The fourth theme is the interplay between international, regional, and national governance structures and standards, and it raises the question of how, over the longer term, platform governance can either strengthen or undermine

transparent, accountable, and human rights-based content governance in the region. A key example would be the African Commission on Human and People's Rights Declaration of Principles on Freedom of Expression of 2019, a document intended to guide the behaviour of states region at country level. It was developed through extensive consultation with African media, internet and human rights organisations across the continent. It articulates international human rights standards that apply to the internet in from an African perspective for the African context. By bypassing this regional instrument, social media councils undermine African efforts to promote respect for human rights online.

What emerges from the papers is that existing national and regional institutions and governance arrangements (including self-governance and co-governance) can serve as a source for greater platform democracy, and that making use of these can serve to strengthen them. Ignoring them can result in consolidating the counter-productive notions that democracy and human rights are "un-African".

Rethinking Platform democracy in Africa: A Nigerian case study

Khadijah El-Usman
PARADIGM INITIATIVE, LAGOS, NIGERIA

This paper explores how one determines legitimacy and differentiates it from democratic platform governance, which standards platforms can adopt in terms of governance to make themselves democratic and legitimate, especially in decision making, rule enforcing and decision reviewing. It further explores whether rule enforcing and decision reviewing be community-based, national-based or regional, considering the possibility of authoritarian governments trampling on rights. The paper analyses these questions from a Nigerian perspective.

Introduction

In 2020, Nigerian youth took to the street to protest police brutality after a series of reports emerged of extrajudicial killings orchestrated by a special unit of the police force.[1] This protest was organised around the country using Twitter. Via this platform, the youth mobilised, canvassed for and distributed supplies and donations while also spreading awareness in a manner that's been unprecedented since the nation's return to democracy.[2] It was a pivotal moment as platforms, especially social media, have become synonymous with providing individuals with unprecedented self-determination through freedom of expression, participation in the democratic process and its public square nature of discussion and information dissemination. During this time, credible news sources quoted social media posts and live streams, many of said news channels were eventually fined for doing so, basing entire news stories on decentralised reports from users of these digital channels.[3]

More upheavals followed this digital revolution. In early 2021, the president of Nigeria, via an official account, put out a tweet that was mass reported for being abusive and tribalist. Twitter ultimately took the tweet down for failing to adhere to its content moderation rules. Consequently, there was an uproar originating from members of government and the ruling administration at the time. Despite a considerable number of cheers from Nigerian users of the microblogging site, the actions of Twitter's leadership were seen by some as an affront to Nigeria's sovereignty. Many have since linked these events to the infamously named "Twitter ban" in Nigeria, which ran from June 2021 to January 2022.[4]

These events have since become catalysts for Nigerians to look deeper into the questions of platform legitimacy and democracy. They've also necessitated questions about the standards to be adopted by platforms in decision-making and what models will work best. This paper discusses considerations

[1] Chiamaka Ozulumba, What Led to #EndSARS Protests?, Thisdaylive, https://www.thisdaylive.com/index.php/2021/10/20/what-led-to-endsars-protests/#:~:text=EndSARS%20protests%20began%20as%20a,citizens%20and%20violated%20their%20rights.Accessed 25th February, 2023

[2] Emmanuel Elebeke, #ENDSARS Coverage: Why we fined AIT, Channels, Arise TV, N3m each — NBC. Vanguard

[3]https://www.vanguardngr.com/2020/10/endsars-coverage-why-we-fined-ait-channels-arise-tv-n3m-each-%E2%80%95-nbc/#:~:text=They%20include%3A%20Channels%20Television%2C%20African,General%20of%20the%20Commission%2C%20Prof Accessed 25th February, 2023

[4] Aljazera, Nigeria ends its Twitter ban after seven months, https://www.aljazeera.com/economy/2022/1/12/nigeria-ends-its-twitter-ban-after-seven-months#:~:text=Nigeria%20ends%20its%20Twitter%20ban,Social%20Media%20News%20%7C%20Al%20Jazeera, Accessed 25th march,2023

revolving around platform legitimacy among Nigerian citizens and how best platform democracy can be achieved; it further explores stakeholders to be involved in platform decision-making and the geographical scope of these decisions.

Defining legitimacy/democracy

Platform legitimacy and democracy, although related, are separate concepts.

Legitimacy refers to the recognition and acceptance of authority or power by the people subject to that authority.[5] It is based on the idea that the exercise of power should operate through consent and be seen as fair and just by those subject to it.

Democracy, on the other hand, is a form of government in which power is held by the people, either directly or through elected representatives. It is based on the idea that the people should have a say in how they are governed and that the government should be accountable to the people.[6]

Relating both concepts to the context of platforms, legitimacy would translate to a platform being wholly accepted by users from its content guidelines, terms and conditions to simply its existence as a part of daily living. Similarly, platform democracy would constitute the extent to which people using this platform are involved in exercising that power and have a say in its governance.

In Nigeria, a certain level of legitimacy has been attained by platforms. The most significant indicator of this has been citizen pushback against the government when it sought control over these platforms or tried to dictate what kind of content should be published and disseminated. In the wake of a Twitter ban, citizens fought and protested at various levels to have the platform restored. Multiple legal cases were instituted across several courts, and many sought to circumvent the block using Virtual Private Networks (VPNs) to access the platform. Most citizens sided with the platform's decision to take down the President's tweet. The action was notably viewed as another avenue for the electorate to exercise its power against the most powerful man in the nation. This show of strength indeed showed the platform had gained some form of legitimacy.

Democracy on platforms

Although platforms have gained acceptance, there is little evidence to suggest they are committed to being democratic. While social media councils - the bodies in charge of platform advice and adjudication with content moderation policies - exist, there is still room for a lot more improvement. While Meta's Oversight Board includes greater diversity than other major social media platforms, it's still not quite democratic. Meta chose the initial board members, and when their terms are up, they will choose their replacements.[7]

Wider questions regarding the ultimate motivations of platform owners and operators remain. With the biggest platforms being largely based outside Africa, their African users have not had a say in how they are governed. Consequently, the nuances of indigenous languages, regional and political contexts, and

[5] Ian Hurd. Legitimacy, https://pesd.princeton.edu/node/516#:~:text=Legitimacy%20is%20commonly%20defined%20in,toward%20the%20rule%20or%20ruler, Accessed 15th march,2023

[6] Stanford Encyclopedia of Philosophy, Democracy, https://plato.stanford.edu/entries/democracy/#DemoDefi, Accessed 15th march,2023

[7] Oversight board bye-laws, https://transparency.fb.com/sr/oversight-board-bylaws-2023, Accessed 11th March, 2023

larger interests of African users (beyond revenue for the platforms and ever-increasing user bases) are often missed.

The emergence of these platforms, in disrupting the distribution of power and information in varying degrees, poses a novel set of challenges to state authority. It affects state sovereignty and in response, the Nigerian government has sought to make many attempts to control this space. The current administration has attempted to institute various forms of restrictive legislation, including a hate speech bill, a social media bill and a code of practice for computer service platforms. Each has been put forward in a bid to wield powers that would ensure a certain style of ownership of the decision making processes of the platforms. Some areas of concern have been mis/disinformation, hate speech, unlawful content, and content taking down time among others. Although these seem on the face of it to be beneficial interventions for citizens, the bills often have provisions that can be subject to abuse by government. One of such provisions can be found in the Hate Speech Prohibition Bill 2019[8] which states that anyone found guilty of hate speech is liable to 3-5 year imprisonment while the social media bill[9] was to allow the government to direct the internet to be shut down, restricting the rights to freedom of expression and access to information. The provisions do not come off as commensurate to the crime, and many feared they would be used to target any government dissidents, as has been witnessed with the Cybercrimes Act 2015[10] where journalists were arrested for reporting unfavourable against the government of the day. Although certain provisions in these bills could be seen as positive, any user of the Internet could fall on the wrong side of the proposed bill, as "falsehood" could have relative meanings.

These attempts were summarily condemned and protested[11] against by distrusting citizens especially because the synergy sought for between the platform and the state seemed to propose attention being paid to only government interests.

After the Twitter ban, the government presented Twitter with a raft of terms including being registered in Nigeria, opening a Nigerian office among others, citing them as conditions that would favour the people and facilitate its return. Of the conditions given by the Government, providing a direct channel to the Partner Support Portal for law enforcement, government officials and Twitter staff to manage prohibited content that violates Twitter community rules was met with the most distrust due to established patterns.

Where the government seeks to be involved or enforce its own platform governance rules, it is often conflicting with national and international human rights principles. Such demands far too often aim to fit a political agenda. On the other hand, when platforms enforce their decision making process or act slowly on certain reports, it often lacks indigenous contexts such as missing hate speech or abusive language said in indigenous language. Therefore, both outcomes leave considerable possibilities for harm to be perpetuated.

[8]Hate speech (prohibition) bill 2019, HB 246, https://placbillstrack.org/upload/HB246.pdf, Accessed 11th March, 2023

[9] Protection from Internet Falsehood and Manipulation Bill 2019, https://guardian.ng/wp-content/uploads/2019/11/Protection-from-Internet-Falsehood-and-Manipulation-Bill-2019.pdf, Accessed 9th March, 2023

[10] Premium Times. Journalist arrested, charged under cybercrime law in Nigeria, https://www.premiumtimesng.com/news/top-news/318321-journalist-arrested-charged-under-cybercrime-law-in-nigeria.html?tztc=1, Accessed 13th March, 2023

[11] John Akubo, Tina Abeku (Abuja), Opeyemi Babalola (Lagos) and Akin Alofetekun (Minna), Protests against hate speech, social media bills rock Abuja, Lagos, https://guardian.ng/news/protests-against-hate-speech-social-media-bills-rock-abuja-lagos/, Accessed 10th March, 2023

Stakeholder involvement in decision making

Social media councils have gained popularity with platforms.[12] They are instituted with the aim of serving advisory and judicial roles in decision making. However, a few flaws are rife in their practice with regard to Africa. First, there is the fundamental flaw of attempting to use one brush to paint all canvases, which is to say that using one global council often based in Europe or America, is unsuitable. Second, is the self regulatory, voluntary implementation model of most social media councils[13] and that existing Social Media Councils lack meaningful participation by African users/citizens.

Establishing a regional or national Social Media Council with stakeholder participation could be a solution and entrenching such a practice will assist platforms to attain increased legitimacy among the people or users. There are however various demographics of stakeholders to involve. Government agencies in Nigeria have adopted a model of stakeholder participation in decision making, that could be used as role model for platform governance. Various demographics and bodies are often represented on the agency's governing board. [14] One of such agencies of note is the National Information Technology Development Agency (NITDA), whose board is statutorily mandated to consist of government agencies in ICT, representatives of computer service bodies, academic staff, representatives from six geopolitical zones and others. It is, however, often criticised for not reaching all the necessary stakeholders.[15] A Nigerian-national based social media council should have the compulsory addition of the human rights commission, civil society organisations and youth participation. This consideration will take advantage of the massive youth participation on platforms, ensuring that a critical mass of users have their concerns and interests represented via relevant and inclusive governance.

National and regional platform governance: the ECOWAS and OHADA models

The question of whether a national or regional approach should be adopted in platform regulation in Africa is a complex one. It is easily made more complicated when viewed in the light of the operation of social media councils. Arguments can be laid out for both approaches.

A national approach to platform governance could allow for more targeted and situation-specific rules that take into account each country's unique needs and challenges. It could also encourage people to take more ownership of and part in making policies and practices, which could help build trust and legitimacy in the regulatory process.

On the other hand, a regional approach to platform regulation could promote greater harmonisation and consistency in the regulatory landscape across different countries in Africa. This could help reduce confusion and fragmentation, and also provide a more level playing field for platform companies and

[12] Article 19, Social Media Councils: One piece in the puzzle of content moderation, https://www.article19.org/resources/social-media-councils-moderation/, Accessed 10th March, 2023

[13] Riku Neuvonen, Internet and Self-Regulation: Media Councils as Models for Social Media Councils?, https://graphite.page/gdhrnet-platform-response/assets/documents/GDHRNet-BestPractice-PartV-2.pdf, Accessed 5th March, 2023

[14] National Information Technology Development Agency Act 2007 Act No. 28 Published In The Federal Republic Of Nigeria Official Gazette No. 99 Vol. 94 Lagos 5th October 2007, https://nitda.gov.ng/wp-content/uploads/2020/11/NITDA-ACT-2007-2019-Edition1.pdf, Accessed 3rd March 2023

[15] Sultan Quadri, Nigeria's House of Assembly adjourns public hearing on controversial NITDA bill, https://techcabal.com/2022/12/24/nigerias-house-of-assembly-adjourns-public-hearing-on-controversial-nitda-bill/, Accessed February 29th 2023

users across the region. A regional approach could also facilitate greater collaboration and information sharing among clusters of countries in addressing common challenges and issues related to social media.

An argument significantly in favour of the regional approach is the propensity of African governments to become authoritarian. For instance, in 2021, 12 countries barred internet access for their citizens at least 19 times on the continent.[16] Thus, where a government becomes authoritarian or espouses authoritarian principles, platforms must return to the role of respecting human rights with the regional body being able to override the national body.

Yet, Africa's complexities, as exhibited by its regional classifications from geography to language and norms, could pose a challenge. This diverse range of interests could make it harder for the entire continent to agree on one form of platform regulation. Meeting halfway, Africa's regional economic communities could serve as orientation, and in Nigeria's case, this would be the Economic Community of West African States (ECOWAS)[17].

The ECOWAS comprises three arms of governance, namely, the Executive, the Legislature and the Judiciary. All three arms in concert, and within them is an established structure that sees to the application of Community laws, protocols and conventions. ECOWAS has unfortunately struggled with adopting and implementing its own laws within the region because although ECOWAS laws are adoptable at a regional level, most member states will need it to be integrated into national laws by their own parliaments.[18] This provision is also in the ECOWAS Revised treaty[19] which states that

"Each Member State shall, in accordance with its constitutional procedures, take all necessary measures to ensure the enactment and dissemination of such legislative and statutory texts as may be necessary for the implementation of the provisions of this Treaty."

This is a difficult process to complete as ECOWAS states often make an argument for state sovereignty. This extends beyond ECOWAS laws to every other treaty. An example of this is the ruling of a Nigerian court in the case Abacha v Fawehinmi, which examined the status of the African Charter on Human and Peoples´ Rights that was domesticated in Nigeria by legislation. The Court held that the Nigerian Constitution is superior to the Charter. However, that is not to say that none of the treaties have managed to be adopted and integrated into national law, essentially practising a dualist[20] system of international law.

A regional body that has managed to achieve this within Africa is the Organization for the Harmonization of Business Law in Africa (OHADA)[21], an intergovernmental organisation for legal integration. It was established by the Treaty of 17 October 1993 signed in Port Louis (Mauritius), as

[16] Access now, Internet shutdowns in 2021 report: resistance in the face of blackouts in Africa, https://www.accessnow.org/internet-shutdowns-africa-keepiton-2021/, Accessed February 20th, 2023

[17] ECOWAS, member states, https://ecowas.int/member-states/, Accessed 27th March, 2023

[18] sec 1(2) of the Constitution of Ghana, see also sec 12(1) of the 1999 Constitution of the Federal Republic of Nigeria.

[19] Article 5(2), Economic Community of West African States (ECOWAS), Revised Treaty of the Economic Community of West African States (ECOWAS), 24 July 1993, available at: https://www.refworld.org/docid/492182d92.html [accessed 30 March 2023]

[20] Carolyn A. Dubay, General Principles of International Law: Monism and Dualism,International Judicial Academy, Washington, D.C., http://www.judicialmonitor.org/archive_winter2014/generalprinciples.html#:~:text=Under%20a%20dualist%20model%2C%20there,and%20its%20citizens%20or%20subjects., Accessed 29th March, 2023

[21] Organisation for the Harmonization of Business Law in Africa,General Overview, https://www.ohada.org/en/general-overview/, Accessed 29th March 2023.

revised on 17 October 2008 in Quebec (Canada) by francophone African countries which Nigeria is not a part of.

What makes the model unique is the practice of the monist[22] system of international law seen in the provisions of Article 10 within it, the Uniforms Acts, which constitute the Ohada Laws, are directly applicable and binding on member states. The Uniform Acts prevail in the event of conflicting provisions of any domestic laws of member-states.

Article 10 states as follows:

"Uniform Acts are directly applicable and overriding in the contracting states notwithstanding any conflict they may give rise to in respect of previous or subsequent enactment of municipal laws."

The Ohada Treaty and laws offer a unique model for the harmonisation of business laws as a starting point for the harmonisation of other laws within ECOWAS states. The Ohada Initiative has created a body of harmonised business laws in its member-states, which have, in consequence facilitated business transactions to a large extent. A similar adoption with ECOWAS or its member states joining OHADA in a bid to implement social media regulation will facilitate platform legitimacy, democracy, and trust within the region. This will not be the first time it has been suggested that ECOWAS adopts the OHADA model or simply joins it. Various authors[23] have posited the same citing the merits of such a merger.

In considering both approaches to platform regulation, what will work best will be a tiered approach. This is because, in practice, both national and regional approaches are likely to be needed in order to effectively regulate platforms in Africa. Countries in the region face unique challenges and circumstances that require tailored and responsive regulatory frameworks while also requiring regional cooperation and coordination in order to address transnational issues such as hate speech, disinformation, and online harassment.

Within a tiered approach, the national body (social media council) is the immediate body for decision making while the regional body will exist as a decision reviewing body. This can be accomplished through the adoption of a uniform legal framework by the new established platform regional body that is based on principles of transparency, accountability, and respect for human rights and must be developed in consultation with a wide range of stakeholders from across the region. States have been known to come together when their interests align, such as in the case of the African Continental Free Trade Area[24], and platform regulation might be one of those places. One key word that is commonly used in both Ohada and Ecowas Treaties is 'integration'. Integration has many facets, including, economic, political, social, geographical, and legal. Any kind of integration must be predicated on a legal framework.

---

[22] n17.

[23] http://www.nigerianlawguru.com/articles/international%20law/OHADA%20TREATY%20AND%20LAW%20BY%20ANGLOPHONE%20STATES.pdf

[24] Tralac,Status of AfCFTA Ratification, https://www.tralac.org/resources/infographic/13795-status-of-afcfta-ratification.html, Accessed 30th March, 2023

Recommendations

User participation in platform decision making and norm setting becomes a priority for social media platforms, taking into consideration national and regional contexts long before establishing social media councils.

In establishing social media councils the adoption of a two tiered system of national and regional platform regulation where the national system covers national decision making and regional body covers decision reviewing should be considered. A tiered approach should be adopted using a regional legal framework for platform decision making and accountability, where a national body exists on decision making, while the regional body takes on decision reviewing using the monist system of international law, where decisions of the regional body surpass that of the national one.

The composition of the national and regional social media councils should comprise various stakeholders and have to include national human rights organisations, civil society organisations and youth representatives.

The overarching aim of this paper was to identify the contextual factors in Nigeria's platform democracy and suggest solutions. Although solutions have been recommended, openng these assertions to further research and testing the models for social media councils practically using case studies, feasibility studies etc. is needed.

The Oversight Board's decisions in the African context

Nashilongo Gervasius Nakale
NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY, WINDHUK, NAMIBIA

The question of self-regulation of platforms through social media councils (many of whom are funded by social media companies) has made the rounds since Facebook (now Meta) announced the establishment of what is called the Oversight Board in 2018. This scenario of a global body making decisions on content takedowns at the national level raises the further question of how self-regulation operates in the context of hyper-globalization, and how this relates to democratic governance and regulation, which has different shapes in different states. Hyperglobalization can be described as "the dramatic change in the size, scope, and velocity of globalization that began in the late 1990s and that continues into the beginning of the 21st century. It covers all three main dimensions of economic globalization, cultural globalization, and political globalization."[1] Its impacts can make it difficult for a government or a country to develop democratic governance on its own terms, maintaining its national sovereignty while also achieving deeper international integration in a hyper-connected world. This is particularly challenging in sub-Saharan countries where digitalization can be linked to new forms of (de)colonisation[2] with nations struggling to maintain self-determination, local traditions, cultures, and values. In an increasingly globalized world universal values are prioritized over local values, particularly online.

Social media councils like the Meta Oversight Board serve as the final layer in the content moderation process, reviewing contested decisions made by the platform and supposedly offering a fair hearing to those users whose content or concerns are affected.[3] The legitimacy of social media councils such as the Oversight Board (OB) is established through the expertise and diversity of its members and the independence of the review process and the resulting decisions. The journey taken by the OB to arrive at either overturning or maintaining decisions made at the platform level involves a process that is supposed to consider local contexts and international human rights standards.

Looking at OB decisions on cases from sub-Saharan Africa during 2021 and 2022 this study found that in practice, their processes pay scant attention to local contexts, values, and regulatory systems. In fact, their decisions appear to almost exclusively reflect globalized values and approaches to online discourse and democracy. In the OB's review of cases, this "globalized" frame of reference takes precedence over efforts to understand the dynamics of the social environments in which the cases take place. What is even more striking is that they do not make reference to regional African human rights frameworks.

While the OB endeavors to be independent in its decision-making process, it's important to understand the challenge it faces from a theoretical perspective. One way of thinking of this challenge is to draw on the "Internet-Governance Impossibility Theorem" which proposes a political Trilemma for the World

[1] https://en.wikipedia.org/wiki/Hyper-globalization

[2] Danielle Coleman, Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws, 24 MICH. J. RACE & L. 417(2019). Available at: https://repository.law.umich.edu/mjrl/vol24/iss2/6

[3] Oversight Board Bylaws. (2022). Oversight Board. https://www.oversightboard.com/sr/governance/bylaws

Economy: one cannot have hyper-globalization, democracy, and national self-determination all at once (Rodrik, 2010 p.200 as cited by Haggart, 2020 p.326).[4] Framed within the role of social media councils and their decisions, this approach states that global interconnectedness leads to limitations of democratic and self-determined solutions. A figurative example would be that a country cannot attain decisions reflecting its sovereignty based on local values including cultural and political nuances, given the digital and political interconnectedness of the world. Within the context of reviewing contested decisions, this would mean that decisions by global platforms on local matters might not necessarily be democratic nor reflective of the sovereignty of a particular state.

The Sub-Saharan African region is a region where democracy presents itself in different ways and degrees, given its diverse governance and political systems, some of which still reflect the region's colonial past.[5] Given that social power structures often mirror the state's, there is a strong likelihood that "universal" or global values might conflict with local ones. An example would be in the case of the 2022 scenario in Uganda where decisions on content based on local anti-homosexuality laws would be irreconcilable with decisions on legitimate content through content moderation based on universal values. If the OB gives precedence to these universal values, this would inevitably result in local laws being disregarded. The research done for this paper suggests this happens more often than not. The result is that the OB is not taking into account the 2019 letter[6] of the Special Rapporteur on Freedom of Expression to Facebook Founder Mark Zuckerberg that highlighted that while international human rights law would provide the board "with a set of tools and a common vocabulary for addressing and resolving hard questions around the moderation of online content" they cannot be used to solve cases as international human rights laws are originally designed to govern the relationship of state authorities with individuals and groups. In this case, the OB is neither a state authority nor does it act on behalf of the state.

This paper looks at decisions on African cases by the Meta OB over the last two years. It examines the process followed, the participation of local people and entities, and the explanations given for the decisions. The paper studies these decisions from a self-regulation and knowledge structure perspective using the Mansell & Steinmueller (2020)[7] framing that self-regulatory practices of digital platforms have the potential to make private decisions on behalf of the public. This framework is worth considering given that the Meta Oversight Board reviews decisions by the platforms that established it, even if the decisions are presented as being 'public'. Hence this paper explores the board's intricate decision-making practices and its potential to make decisions that are private in their nature on behalf of the public, a challenge highlighted by Haggart (2020) who is inclined that there are "specific implications of requiring private corporations to adhere to human rights provisions."

[4] Haggart B (2020) "Global platform governance and the internet-governance impossibility theorem" Journal of Digital Media & Policy, Volume 11, Issue Regulating Digital Platform Power, Nov 2020, p. 321 - 339

[5] Njoh, A. J. (2000). The Impact of Colonial Heritage on Development in Sub-Saharan Africa. Social Indicators Research, 52(2), 161–178. http://www.jstor.org/stable/27522501

[6] D.Kaye, (2019) Letter to Facebook CEO, "Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression" available at https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Legislation/OL_OTH_01_05_19.pdf

[7] Mansell, Robin & Steinmueller, W.. (2020). ADVANCED INTRODUCTION TO PLATFORM ECONOMICS.

Oversight Board decisions on cases from Africa during 2021 and 2022

Decisions brought to the OB are published on the board's website. These are accessible in full text on the body's website depicting the type, location, the date it was published, as well as the country where the case was raised. Of the 35 decisions published by February 2023, only seven of them originated in the African region. Wong and Floridi (2022) highlight this lack of geographic diversity in case reporting from the global South when they point out that "significant geographic regions such as Sub-Saharan Africa and Central and South Asia represent[ing] just two percent of appeals".[8] Africa cases account for only about 1% of the total number of cases. This reflects little knowledge of or access to the appeal processes by users in the region.

The table below briefly highlights the decisions to remove content and their nature.

Description	Nature	Country	Region	Published	Decision
Video After Nigeria church attack	Violent and graphic content	Nigeria	SSA	Dec, 2022	Overturned
Tigray Communication Affairs Bureau	Violence and incitement	Ethiopia	SSA	Oct 4, 2022	Upheld
Reclaiming Arabic Words	Hate speech	Morocco, Egypt, Lebanon	MENA	June 13, 2022	Overturned
Sudan Graphic Video	Violent and graphic content	Sudan	SSA	June 13, 2022	Upheld
Alleged Crimes in Raya Kobo	Hate speech	Ethiopia	SSA	Dec, 2021	Upheld
South Africa Slurs	Hate Speech	South Africa	SSA	Sept 28 2021	Upheld
Shared Al Jazeera Post	Dangerous individuals & organizations	Israel, Egypt	MENA	Sept 14, 2021	Overturned

Three of the above seven decisions on African cases were overturned by the OB, while four of the decisions were upheld. Six of the seven cases stem from countries where conflict exists and/or human rights are constantly under threat. Out of the seven decisions listed in the table, five) are from Sub-Saharan African (SSA) countries, and it is these that the rest of the report concentrates on.

The investigation process

The OB upholds or overturns Meta's decisions. In doing so, the Board considers Meta's content policies and international human rights standards, particularly the International Covenant on Civil and Political Rights. This international/universal approach holds potential challenges according to Barata (2022) who opines that "international human rights law is originally designed to govern the relationship of State authorities with individuals and groups". The application of international laws in local contexts could pose further challenges according to Evelyn Douek (2021) who advocates that "International human rights laws… in content moderation should rather serve the interests of users and society rather than being co-opted by platforms to their own ends"[9]. This scenario proves the applicability of the Internet Governance Impossibility Theorem, which calls for consensus building in solving local issues at the local level, in order to achieve self -determination in democracy.

[8] Wong, D., Floridi, L. Meta's Oversight Board: A Review and Critical Assessment. Minds & Machines (2022). https://doi.org/10.1007/s11023-022-09613-x

[9] Douek, Evelyn, The Limits of International Law in Content Moderation (October 12, 2020). 6 UCI J. INT'L TRAN'L & COMP. L. 37 (2021), Available at SSRN: https://ssrn.com/abstract=3709566 or http://dx.doi.org/10.2139/ssrn.3709566

As part of the Oversight Board's procedures, an explanation and disclaimer are published alongside the cases on how decisions are prepared, the approval process, and what the decisions represent. A review of these explanations and disclaimers done for this paper indicates that, for all the African cases, the OB commissioned independent research by an unnamed institute, headquartered at the University of Gothenburg in Sweden. This institutes's role is to review and substantiate the OB's findings by looking at the local context. A disclaimer on each case reviewed further states that this institute draws on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The OB is further "assisted by an advisory firm focusing on the intersection of geopolitics, trust, and safety, and technology called Duco Advisors, and Memetica, a digital investigation group providing risk advisory and threat intelligence services to mitigate online harms".

Details of the investigation processes for the five cases of focus are highlighted below. The analysis starts with the only case where a decision was overturned. The next table analyses cases where decisions were upheld. The tables include the decisions, the structure of investigations, public comments/third-party submissions as well as policy advisory statements.

Overturned cases

Case	Video after Nigeria church attack
Decision	The Board has overturned Meta's decision to remove a video from Instagram showing the aftermath of a terrorist attack in Nigeria.
Structure of Investigation	1. The user explained that they shared the video to raise awareness of the attack and to let the world know what was happening in Nigeria. 2. The Board asked Meta 29 questions, 28 of which were answered fully. Meta was unable to answer a question on the percentage of user reports that are closed without review in the Sub-Saharan Africa market. 3. The Board looked at the question of whether this content should be restored through three lenses: Meta's content policies, the company's values, and its human rights responsibilities. 4. The Board focused on three of Meta's content policies: Violent and Graphic Content; Bullying and Harassment, and Dangerous Individuals and Organizations. 5. The majority of the Board found that no policy was violated.
Public Comments/Third Party Submissions	Nine public comments related to this case are considered. One of the comments was submitted from Asia Pacific and Oceania, one from Central and South Asia, one from the Middle East and North Africa, one from Sub-Saharan Africa, and five from the United States and Canada. The submissions covered themes including the need to clarify the Violent and Graphic Content policy, and Nigeria-specific issues that the Board should be aware of while deciding this case.
Policy Advisory Statement	Content policy: Meta should review the public facing language in the Violent and Graphic Content policy to ensure that it is better aligned with the company's internal guidance on how the policy is to be enforced. Enforcement: Meta should notify Instagram users when a warning screen is applied to their content and provide the specific policy rationale for doing so.

Upheld cases

Case	Tigray Communication Affairs Bureau
Nature:	Violence and Incitement. Decision to remove a post threatening violence in the conflict in Ethiopia
Structure of Investigation	1. Following Meta's referral and the Board's decision to accept the case, the user was sent a message notifying them of the Board's review and providing them with an opportunity to submit a statement to the Board. The user did not submit a statement. 2. The Board asked Meta 20 questions. Meta answered 14 questions fully and six questions partially. The partial responses related to the company's approach to content moderation in armed conflict situations, imposing account restrictions for violations of content policies, and the cross-check process.

Public Comments:	The Oversight Board received and considered seven public comments related to this case. One of the comments was submitted from Asia Pacific and Oceania, three from Europe, one from Sub-Saharan African and two From the United States and Canada. The submissions covered the following themes: – the inconsistency of Meta's approach in the context of different armed conflicts; -the heightened risk accompanying credible threats of violence between parties during an armed conflict; – the problems with Meta's content moderation in Ethiopia and the role of social media in closed information environments; – factual background to the conflict in Ethiopia, including the harm suffered by Tigrayan people and the role of hate speech against Tigrayans on Facebook in spreading violence; – and the need to consider laws of armed conflict in devising policies for moderating speech during an armed conflict.
Oversight Board Decision	The Oversight Board upholds Meta's decision to remove the content for violating the Violence and Incitement Community Standard.
Policy Advisory Statement	Transparency: In line with the Board's recommendation in the "Former President Trump's Suspension" as reiterated in the "Sudan Graphic Video" Meta should publish information on its Crisis Policy Protocol. The Board will consider this recommendation implemented when information on the Crisis Policy Protocol is available in the Transparency Center, within six months of this decision being published, as a separate policy in the Transparency Center in addition to the Public Policy Forum slide deck. Enforcement: To improve enforcement of its content policies during periods of armed conflict, Meta should assess the feasibility of establishing a sustained internal mechanism that provides the expertise, capacity and coordination required to review and respond to content effectively for the duration of a conflict. The Board will consider this recommendation implemented when Meta provides an overview of the feasibility of a sustained internal mechanism to the Board.

Case	**Sudan graphic video**
Nature:	Violent and graphic content. Decision to restore a Facebook post depicting violence against a civilian in Sudan.
Structure of Investigation	1. Following Meta's referral and the Board's decision to accept the case, the user was sent a message notifying them of the Board's review and providing them with an opportunity to submit a statement to the Board. The user did not submit a statement. 2. The Board asked Meta 21 questions. Meta responded to 17 fully and four partially. The partial responses were to do with questions on measuring the impact of Meta's automated system on content on the platform and why the Violent and Graphic Content Community Standard does not contain a raising awareness exception.
Public Comments:	The Board received five public comments for this case. Two comments were from Europe, one from Sub-Saharan Africa, and two from the United States and Canada. The submissions covered the following themes: the need to adopt a more context-sensitive approach that would set a higher threshold for removal of content in regions subject to armed conflicts, so that less content is removed; the need to preserve materials for potential future investigations or to hold violators of human rights accountable; and that the newsworthiness allowance is likely to be applied in an ad hoc and contestable manner and that this practice should be reconsidered. In March 2022, as part of ongoing stakeholder engagement, the Board spoke with approximately 50 advocacy organization representatives and individuals working on reporting and documenting human rights abuses, academics researching ethics, human rights, and documentation, and stakeholders interested in engaging with the Board on issues arising from the Violent and Graphic Content Community Standard and its enforcement in crisis or protest contexts.
Oversight Board Decision	The Oversight Board upholds Meta's decision to leave up the content with a screen that restricts access to those over 18.
Policy Advisory Statement	Content policy: Meta should amend the Violent and Graphic Content Community Standard to allow videos of people or dead bodies when shared for the purpose of raising awareness of or documenting human rights abuses. This content should be allowed with a warning screen so that people are aware that content may be disturbing. The Board will consider this recommendation implemented when Meta updates the Community Standard. Enforcement: To ensure users understand the rules, Meta should notify users when it takes action on their content based on the newsworthiness allowance including the restoration of content or application of a warning screen.

Case	South Africa slurs
Nature:	Hate speech. Decision to remove a post discussing South African society under its Hate Speech Community Standard.
Structure of Investigation	1.Facebook notified the user that their post violated Facebook's Hate Speech Community Standard. Facebook stated that the notice to the user explained that this Standard prohibits, for example, hateful language, slurs, and claims about the coronavirus. 2.The user appealed the decision to Facebook, and, following a second review by a moderator, Facebook confirmed the post was violating. The user then submitted an appeal to the Oversight Board. 3.The Board asked Facebook how its market-specific slur list is enforced, and if a slur's appearance on any market list means it cannot be used globally. Facebook responded that its "prohibition against slurs is global, but the designation of slurs is market-specific, as Facebook recognizes that cultural and linguistic variations mean that words that are slurs in some places may not be in others." The Board reiterated its initial question. Facebook then responded "[i]f a term appears on a market slur list, the hate speech policy prohibits its use in that market.
Public Comments:	The Oversight Board received six public comments related to this case. Three of the comments were from Sub-Saharan Africa, specifically South Africa, one was from Middle East and North Africa, one was from Asia Pacific and Oceania, and one was from the United States and Canada. The Board received comments from stakeholders including academia and civil society organizations focusing on freedom of expression and hate speech in South Africa. The submissions covered themes including the analysis of the words "clever blacks," "n***er" and "k***ir;" whether the words "n***er" and "k***ir" qualify as hate speech; the user's and reporter's identity and its impact on how the post was perceived; and the applicability of Facebook's Hate Speech policy exceptions.
Oversight Board Decision	The Oversight Board upholds Facebook's decision to remove the content.
Policy Advisory Statement	Facebook should: Notify users of the specific rule within the Hate Speech Community Standard that has been violated in the language in which they use Facebook, as recommended in case decision 2020-003-FB-UA (Armenians in Azerbaijan) and case decision 2021-002-FB-UA (Depiction of Zwarte Piet). In this case, for example, the user should have been notified they violated the slurs prohibition.

Case	Alleged crimes in Raya Kobo
Nature:	Hate Speech. Decision to remove a post alleging the involvement of ethnic Tigrayan civilians in atrocities in Ethiopia's Amhara region.
Structure of Investigation	1. Meta notified the user that his post violated Facebook's Hate Speech Community Standard, but not the specific rule that was violated. The user then appealed the decision to Meta, and, following a second review by another moderator from the Amharic content review team, Meta confirmed that the post violated Facebook's policies. 2. The user then submitted an appeal to the Oversight Board. As a result of the Board selecting the case, Meta identified the post's removal as an "enforcement error" and restored it on August 27. Meta stated that it usually notifies users about content restoration on the same day. However, due to a human error, Meta informed this user of restoration on September 30.
Public Comments:	The Oversight Board received 23 public comments related to this case. Six of the comments were from Sub-Saharan Africa, specifically Ethiopia, one from the Middle East and North Africa, one was from Asia Pacific and Oceania, five were from Europe and ten were from the United States and Canada. The Board received comments from stakeholders including academia, private individuals and civil society organizations focusing on freedom of expression and hate speech in Ethiopia. The submissions covered themes including whether the content should stay on the platform, difficulties in distinguishing criticism of the TPLF from hate speech against the Tigrayan people, and Meta's lack of content moderators who speak Ethiopian languages.
Oversight Board Decision	The Oversight Board upholds Meta's original decision to remove the content. Given that Meta subsequently restored the content after the user's appeal to the Board, it must now remove the content once again from the platform.
Policy Advisory Statement	Meta should: rewrite Meta's value of "Safety" to reflect that online speech may pose risk to the physical security of persons and the right to life, in addition to the risks of intimidation, exclusion and silencing.

	Facebook's Community Standards should reflect that in the contexts of war and violent conflict, unverified rumors pose higher risk to the rights of life and security of persons. This should be reflected at all levels of the moderation process. Meta should commission an independent human rights due diligence assessment on how Facebook and Instagram have been used to spread hate speech and unverified rumors that heighten the risk of violence in Ethiopia. The assessment should review the success of measures Meta took to prevent the misuse of its products and services in Ethiopia.

Looking at this table reveals that the appeal system mostly receives input from ourside sub-Saharan Africa. This raises the question of access to the appeal system by people from sub-Saharan Africa, particularly the subcommunities in the country of origin of the case who might be affected by the content.

OB decisions and national democratic governance and self-determination

To understand how the OB's decisions relate to democratic governance within the country of origin of the complaint it is necessary to look at the political systems in the concerned countries. This approach can also help reveal how democracy at the national level functions in the context of hyper-globalization. As highlighted earlier, all five cases emanate from countries where either war or conflict is currently brewing, and/or where there is a history of racial violence and divides.

- Nigeria, ongoing political and religious tensions, embedded in the tribal, political, and religious fabric of that nation
- South Africa, racial and economic tensions dating to previous colonial governing systems
- Ethiopia has an ongoing regional/ethnic tension, brewing from political governing systems
- Sudan too has ethnic and political tensions stemming from issues of borders and contested leaderships shaped by colonial history

These historical and current political systems shape these society's values, sensitivities and norms, and these are often reflected in the behaviors of citizens, including in their engagements on online platforms. It is also worth noting that most of these countries are signatories to international human rights treaties and do have human rights laws and institutions. By appealing the decisions of the platform, users are indicating that content moderation does not show an understanding of the local context. In more direct terms, the platform's original decisions made based on its rules and guidelines appear to be one-dimensional. This is where the OB is intended to provide a broader perspective and consider local contexts. However, it appears that the decisions of the OB are also inadequate as they are made based primarily on a framing of the issues only based on universal values and laws even if the context requires deeper cultural/social and political sensitivity. This is demonstrated by the lack of direct use of local or regional laws in the cases at hand. By only paying attention to community standards and guidelines and the application of international human rights law, the OB reinforces the argument that platforms can undermine the self-determination of people within a national democratic space.

An alternative approach would have been for the OB to pay more attention to national legislative and human rights frameworks – and even regional Africa frameworks - and to convene more regional consultations. This would complement their existing process. By involving local and regional bodies including NGOs, regulatory bodies, and even digital oversight mechanisms at the African Union level and other relevant regional bodies they can make decisions that are more respectful of national democratic laws and processes.

The situation of applying universal laws and values to local contexts further translates to imposing global laws in the local context and disregarding local and regional guidelines (when in place) and laws and

indirectly enforcing historical narratives of bigger power in local matters. In the cases under focus and informed by their review and decision-making processes, it is clear that the decisions of the OB did not consider or complement local laws and norms. The case of the Tigray Communication Affairs Bureau becomes a key example here. In this case, we see content created from within a specific region where people are disgruntled with the current political situation. Within that context, the content was flagged to be taken down and even though two Amharic-speaking reviewers "determined that the post did not violate Meta's policies and left it on the platform", Facebook through another layer of content review for "high-risk situations" went ahead and removed it. As demonstrated by the decision's explanatory notes, the Board did not consider how the local communications/editorial or media authority would have responded to or viewed this content. Consequently, this can be interpreted to be disrespectful of local contexts while amplifying international norms. By only looking at international human rights frameworks, the OB appears not to consider efforts within these countries to adhere to and implement the international human rights agreements to which they are signatories. This affirms the notion that the self-determination of nations/people is indeed difficult to achieve within hyper-globalization.

Covert ways of the enforcement of universal values and international laws within local contexts are demonstrated by the OB's investigation processes. They are carried out from afar and consultations are undertaken had limited public input from the affected country and the region. In studying the decisions of the OB in SSA, this study did not only find that the majority of the Sub-Saharan African cases received a low number of submissions from the public and third parties in general; it revealed that the majority of submissions were not from countries and regions that share history and political systems with the countries from which the cases came. More public and third-party submissions in support of the users from the relevant country and the region submitted to the OB before it made its final decisions might have presented the OB with a different viewpoint which it would have had to consider in its treatment of these cases.

For example, it might have encouraged the OB to put more effort into looking at the relationship between universal human rights standards and the dominant narratives of the countries of origin of the platforms. This relationship can be complex, made up of a mix of areas where these rights and standards conflict or coincide. This approach would respond to the challenges of current social media councils' regulations as criticized by Flew. et al (2020) that "the challenges of media regulation in the digital age arise from the complexity of regulating in a contested global arena where national policies are often in conflict and laws are not always enforceable in a straight-forward way".[10] In essence, considering national/regional policies and laws first, would add counter perspectives to this global perspective on local context as well as convey that the OB takes local contexts seriously, considering them before switching to applying global frameworks.

The question of self-regulation with regard to the knowledge structure of Meta and the OB remains central as posed by Mansell & Steinmueller (2020) who framed that self-regulatory practices run the risks of digital platforms making private decisions on behalf of the public. The OB makes 'public decisions' based on a public process, but it was set up as a self-regulatory arm of the former Facebook. I.o.w. Mansell's framing applies. The Board with funding from Meta engages consulting agencies for research aimed at informing the decisions they make on cases brought to them. The Board further sends questions to the platform to respond to (it does not always do so). The OB analyses the rules of the

[10] T. Flew & M. Fiona &N. Suzor (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. Journal of Digital Media & Policy. 10. 33-50. 10.1386/jdmp.10.1.33_1.

platform against the complaints of the users. While in some cases the Board seems to be referring to multistakeholder consultations which seem to have happened out of the scope of the cases in this regard (as in South Africa), it does not appear to have considered the media or social media regulations within a specific country. Ideally, this could have happened through consultations with the communications regulators within each country and or engagements with media councils, ombudspersons, and even the local internet governance systems in these countries.

In conclusion, this study finds that: the decision-making model of the Meta Oversight Board as a social media council is not entirely independent given that it omits to work with local policy, regulations, or other bodies, as well as taking global standards into account. It also fails to demonstrate how consultations at the local level are considered within their reviewing process. While it can be assumed that the consulting agencies they use would carry out this work, It is important for this fact to be reflected in their description of the process to avoid perceptions of being detached from local realities, applying instead globalized approaches to local contexts. Previous findings from research such as that by Barata (2020)11 support this conclusion by indicating that "the approach of the Oversight Board has exclusively been based on universal standards" which consequently disregards local standards and laws. While recognizing that the Oversight Board is overseeing decisions made by global platforms and must uphold international laws, this research has indicated that it has failed to apply what is widely understood as the multistakeholder approach to internet governance that uses consensus building in making decisions.

Recommendations

Given the findings of this paper, the following recommendations are proposed:

- The Oversight Board policies need to be explicit and intentional in giving consideration to national and regional laws and norms as the first step in dealing with cases under review
- The Oversight Board should engage directly with multistakeholder processes within the country and regions of their decisions as proposed by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in 2019.
- The Oversight Board should indicate how and with who their research consultants engaged people and institutions (including media regulators and human rights institutions) in the relevant countries and reflect this within their case decisions.
- For the African Region, the Oversight Board should demonstrate its inclusion of engagements with regional laws and regional bodies as a necessity in trusting that their processes are focused on dealing with local contexts. This will ensure transparency and build the legitimacy of regional oversight structures and internet-related human rights norms and standards.
- The Oversight Board should invest heavily in awareness campaigns in Africa about their appeal procedures for cases as well as mechanisms of public input within their investigating processes.

[11] J.Barata "The Decisions of the Oversight Board from the Perspective of International Human Rights Law" Special Collection of the case law on Freedom of Expression, Colombia University, available online at https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2022/10/The-Decisions-of-the-OSB-from-the-Perspective-of-Intl-Human-Rights-Law-Joan-Barata-.pdf

Elections and social media platforms in Kenya

Grace Githaiga
KENYA ICT ACTION NETWORK

Introduction

Elections are scheduled to take place in more than 70 nations in 2024, including some of the largest democracies on earth and others with weak or factually nonexistent democracies. A rising number of African countries with recent ones being Kenya in 2022, and Nigeria in 2023, have recently deployed tech for elections. This has been in the hope of inspiring confidence, increasing transparency and trust, especially in presidential elections that are mostly contested.

These recent elections have also seen a lot of use of social media platforms by politicians to share their manifestos, advance electoral ideals, conduct campaigns, motivate and galvanise their supporters around issues that are deemed of interest to these followers. Notably, in Kenya, some politicians attracted such huge following on such platforms as Meta and Twitter and ended up having ‘‘broadcasting stations”[1] on these platforms, where they would have daily broadcasts targeting their supporters during the campaign season in 2022. In these daily broadcasts, citizens were sometimes exposed to fake news, disinformation, hate speech and fake photos of “large crowds” at campaign rallies.[2]

Candidates in Kenya frantically tried to catch the attention of the nation's 12 million social media users due to the growth in popularity of platforms like Twitter and Facebook.[3] A phenomenon occurred where requests to rent out Facebook pages and Twitter accounts from influencers with hundreds of thousands of followers became rampant, mainly for political purposes. These influencers gave politicians the chance to disseminate their views, address criticism, or even generate rumors about their rivals.[4] Very few candidates could therefore afford to ignore social media. This became good business for the influencers who would get paid a daily rate of 400 euros for a hashtag's development and maintenance.[5]

The two main presidential candidates, Dr. William Ruto of the Kenya Kwanza Alliance, and Mr. Raila Odinga of Azimio la Umoja-One Kenya Coalition Party “invested heavily in building digital war chests to interact with their supporters and market their manifestos and agenda”.[6] Nonetheless, both coalitions suffered from fake news spear campaigns mostly fueled by influencers working for the different coalitions.[7]

---

[1] Kenya’s President Ruto who has a 2.3M following would run daily broadcasts of his campaigns in 2022. https://www.facebook.com/williamsamoei.Another example is former Nairobi Governor Mike Sonko with a following of 2.5M https://www.facebook.com/search/top/?q=Mike%20Sonko and still runs Sonko TV which has 104K follower https://www.facebook.com/MikeSonkoTv001.

[2] Githaiga, Grace. 2022. Disinformation Factories in Kenya. https://www.linkedin.com/posts/gracegithaiga_election-influencers-for-hire-kenyas-disinformation-activity-6928056612398010368-VyzG/?trk=posts_directory.

[3] Caronline Rabut. 2022. Election influencers for hire: Kenya’s disinformation factories. calendar.google.com/calendar/u/0/r?tab=mc

[4] AFP. 2022. Social media Influencers cash in as presidential election approaches in Kenya. https://www.africanews.com/2022/05/05/social-media-influencers-cash-in-as-presidential-election-approaches-in-kenya/.

[5] Ibid.

[6] Steve Omondi. 2022. How social media influencers shaped Kenya’s 2022 General Election, https://mediainnovationnetwork.org/2022/08/29/how-social-media-influencers-shaped-kenyas-2022-general-election/.

[7] Ibid.

A study by the Mozilla Foundation, observed that social media sites such as Facebook, TikTok, and Twitter failed to uphold their promises of election integrity during Kenya's 2022 elections. Further, it was found that political advertising contributed to amplification of propaganda, and the platforms content labeling was unable to combat disinformation.[8] Twitter and Tiktok's labeling efforts were sporadic and ineffective in halting the spread of disinformation. The study further averred that this affected some political parties more negatively especially when these platforms went ahead to announce the election results ahead of the formal announcement, making them appear partisan. As for facebook, the platform was seen as lacking visible labels, which allowed for the dissemination of propaganda. The Mozilla study deduced that platforms therefore failed to live up to their promises to stakeholders to be reliable sources of election information and instead ended up being avenues where rumors, conspiracies, and false information thrived.

Kenya has a population of over 50 million, with half of them being under 35 years. Further, 12 million of them utilize social media.[9] Considering that a large part of the African electorate are young people who are on social media platforms, what ethical concerns does this raise for the platforms? How are platforms responding to these concerns considering some, such as facebook, twitter, tiktok have now become mainstream and especially during electioneering periods as observed in Kenya? What is the level of preparedness of these platforms in terms of content moderation?

This paper briefly responds to some of these issues drawing on Kenya’s examples during the electioneering period of 2022.

Guiding questions

In this section, three guiding questions are addressed.

Ethical Concerns for Platforms

Social media platforms provided unprecedented opportunities to both candidates and their supporters to engage, during the electioneering period in Kenya. There was an increase in the number of politicians making use of social media platforms in 2022 compared to the 2017, and 2013 elections. But with this surge, the platforms were abused to amplify social differences and influence citizens' decision-making processes.[10] The social media space was transformed into a ferocious battleground where bloggers from Kenya’s two main political parties spewed hate speech and used all kinds of dirty techniques to psychologically edge out their rivals in the hotly contested vote.[11] The platforms steadily accumulated virtual "litter" in the form of fake news, misinformation, disinformation, and propaganda from the pre-campaign period through the official three-month campaign period to the election and post-election period.[12]

[8] Odanga, Madung. 2022. Opaque and Overstretched: how platforms failed to curb misinformation during the Kenya 2022 election. https://foundation.mozilla.org/en/campaigns/opaque-and-overstretched-part-ii/#case-study-labeling-failures.

[9] AFP. 2022. Social media Influencers cash in as presidential election approaches in Kenya. https://www.africanews.com/2022/05/05/social-media-influencers-cash-in-as-presidential-election-approaches-in-kenya/.

[10] UNESCO. 2022. How UNESCO works to curb online hate speech and disinformation ahead of Kenyan general elections. https://www.unesco.org/en/articles/how-unesco-works-curb-online-hate-speech-and-disinformation-ahead-kenyan-general-elections.

[11] Steve Omondi. 2022. How social media influencers shaped Kenya’s 2022 General Election. https://mediainnovationnetwork.org/2022/08/29/how-social-media-influencers-shaped-kenyas-2022-general-election/.

[12] Ibid.

There have been efforts by different stakeholders to confront the challenges. For example, the government's response has been through the establishment of the National Cohesion and Integration Commission (NCIC) in 2008.[13] This was done under the NCIC Act No.12 of 2008. The purpose of the NCIC was to create a national institution to advance national identity and values, reduce ethnic-political rivalry and violent crime, end racial, ethnic, and religious discrimination, and foster national healing and reconciliation. In application on social media, Section 62 of the Act outlaws speech that is designed to provoke contempt, hatred, hostility, violence, or discrimination against any individual, group, or community on the basis of ethnicity or race.[14]

Platforms such as Twitter, Meta, and TikTok all indicated that they were battling false information during 2022 Kenya's general elections.[15] They revealed that they had deployed moderators to remove harmful content. However, civil society groups felt that because of these moderators' lack of understanding of regional languages and dialects, and the lack of standards that take into consideration certain cultural and societal settings, bad content was spread swiftly with potentially harmful results.[16]

Several challenges still need attention from platforms. For example, there is the young population in Kenya on social media platforms and who are potential voters come the next general election. The platforms will need to engage this section of the population to raise awareness on responsible use of social media, but also collect their views on ways in which they think the platforms can respond to challenges for example of hate speech and disinformation. This can be done through focus groups, surveys, and in-depth interviews. These are methods for involving people in community development and asking for their perspectives.[17] Another challenge is that of the AI language tool(s) such as ChatGPT[18] that might see a significant amount of seemingly original content being produced by the use of generative language models, relieving propagandists of the need to repurpose the same text across several news websites or social media platforms.[19]

There is therefore a need for platforms to address these ethical concerns in a timely fashion, in order to contain the rise in political tensions using platforms, and especially during the electioneering period.

What are platforms doing- and is it enough?

There is no doubt that tech platforms continue to be used in Kenya to advance electoral ideals and manifestors. But with an increase in use, new challenges continue to emerge such as fake news, disinformation, propaganda, including insults against opponents, threats and incitement to violence,

[13] The National Cohesion and Integration Commission (NCIC). https://cohesion.or.ke/.

[14] RefWorld. 2013. Tackling Online Hate Speech in Kenya. https://www.refworld.org/docid/51308a312.html.

[15] Nita Bhalla. 2022. Call to action: Kenyan elections under threat of hate speech. https://www.csmonitor.com/World/Africa/2022/0627/Call-to-action-Kenyan-elections-under-threat-of-hate-speech.

[16] Ibid.

[17] Behaviorial Insights Team. 2022. How can citizens shape social media platforms? https://www.bi.team/blogs/how-can-citizens-shape-the-future-of-social-media-platforms/.

[18] Start using Chat GPT. https://ai-pro.org/start-chat-gpt/?ppg=03&keyword=chatbot%20ai&adid=647798462831&gclid=Cj0KCQiAgaGgBhC8ARIsAAAyLfF6TOy3RWazp-F3Go_Z8mgNKBRQQGhoUzUfqFdQ8X5VVmR1RbaGeQcaAqU1EALw_wcB

[19] Timcke, Scott . 2023. Generative Language Models in Algorithmic Social Life: Some Concepts and Considerations. Research ICT Africa. https://researchictafrica.net/2023/01/12/generative-language-models-in-algorithmic-social-life-some-concepts-and-considerations/.

defamatory language, propaganda, and covert hate speech etc. Such challenges are contributing to the calls from different stakeholders for content moderation on platforms.

In July 2022, the Kenyan government through the NCIC, indicated that Meta (Facebook) had neglected to fix the problem of hate speech ahead of August's general elections. Using a report produced by Global Witness, there was an indication that English and Swahili, the two official languages of Kenya, were not found to contain hate speech adverts on Facebook.[20] This was not the situation and the issue was that Facebook failed to flag hate speech adverts in either language. Accordingly, the platform was given one week to adhere to laws pertaining to hate speech.[21]

Facebook was not the only culprit. According to research by the Mozilla Foundation, TikTok was seen as a "platform for quick and far-reaching political disinformation".[22] In fact, "[e]ven videos that appear to be in clear violation of TikTok policies seem to have been amplified by the algorithms".[23] In response, Tik Tok removed the videos flagged by the Mozilla Foundation, which the platform then determined to be against its community guidelines.

Twitter indicated that it had employed skilled moderators and worked with independent fact-checkers as part of their aggressive disinformation-fighting strategy.[24] Further, the platform averred its commitment "to protect the health of the electoral conversation." However, it has been felt that the bringing down of offensive posts sometimes takes too long to prevent damage[25].

Engagement of civil society

As part of its civic duty and election work, KICTANet[26] a multistakeholder think tank for ICT policy and regulation and a catalyst for reform in the ICT sector, convened dedicated engagements for civil society actors with both Meta and Twitter. The Network also participated in a meeting that Tik Tok convened for civil society organisations.[27] The dedicated engagements between KICTANet and platforms discoursed on concerns about social media use during elections. It was disclosed that as a way of responding to concerns that had emerged during elections, the three platforms had put in place various mechanisms which aimed at monitoring and addressing false information and hate speech on their platforms. The mechanisms included, among other things, encouraging fact-checking, promoting instructional materials on false information and hate speech, and providing links and election-related information. However, stakeholders felt that more could be done in combating the challenges identified, in order to inspire trust and confidence in users of these platforms. For example, it was expressed that

[20] Samuel Gebre. 2022. Social media platforms under scrutiny ahead of Kenyan elections, https://techxplore.com/news/2022-08-social-media-platforms-scrutiny-kenyan.html.

[21] Ibid.

[22] Ibid.

[23] Andrew Deck. 2022. Hate speech and disinformation spike on TikTok in run-up to Kenya's elections.

https://restofworld.org/2022/hate-speech-and-disinformation-spike-on-tiktok-in-kenya-election-run-up/

[24] Nita Bhalla. 2022. Online disinformation stokes tensions as Kenya elections near. https://www.context.news/digital-rights/online-disinformation-stokes-tensions-as-kenya-elections-near.

[25] My own conclusion after conversing informally with colleagues.

[26] www.kictanet.or.ke. KICTANet's guiding philosophy encourages synergies for ICT policy-related activities and initiatives. The network acts as a catalyst for reform in the ICT sector and is guided by four pillars: policy advocacy, stakeholder engagement, capacity building, and research.

[27] Mwendwa Kivuva. 2022. Emerging Concerns on Social Media use in the upcoming 2022 Elections.

2022, https://www.kictanet.or.ke/emerging-concerns-on-social-media-use-in-the-upcoming-2022-elections/.

the platforms needed to make enough of an investment or take enough action to identify, prohibit, and prevent hate speech, false information, and disinformation on their platforms especially during the electioneering period.[28]

Level of preparedness for content moderation

Social media platforms have found themselves under scrutiny, in particular on how they moderate content during elections. Their level of preparedness has been questioned in particular when it comes to elections in African countries. There have been questions around the criteria they use to moderate content, the tools they use, especially where there exists a multiplicity of local languages as well as dialects.

Organizations monitoring social media and fighting to defend election integrity, noted that social media corporations invest less and provide fewer tools in the Global South than in the North.[29]

There is also now the complex language processing tools like ChatGPT which may have an impact on the operations on social media.[30] It is touted that language models might be able to compete with human-written content for a reasonable price. As such, any potent technology might provide specific advantages to propagandists who choose to use them.[31] It remains to be seen how the platforms will handle these new AI language tools, especially in African elections and whether they will be used to create automated propaganda. It will be interesting to see how the tool for example will write local languages and dialects and the slant in messaging.

Conclusions

Platforms are going to increasingly be used during elections in Africa, mainly as a double sword: to campaign and share political parties' manifestos, sow dissent and spread propaganda of opponents.

In light of this and considering the need to have content moderation policies that might at times need to be context specific, this paper notes that every social media firm has its own policies and procedures regarding online harm, such as with varying definitions of terms like misinformation and disinformation. As such, it would be useful to consider if there is a need to put pressure on social media platforms to adopt uniform definitions for all online lies and disinformation.

There is a lack of clarity on which languages are moderated by the platforms in the different African countries. It would be important to have an idea how this is determined, since Africa has many languages that have different dialects and moderators would need to understand the nuances.

This paper notes that platforms endeavored to engage with stakeholders in Kenya during the electioneering period. This was commendable as engagement with key stakeholders brings about an understanding of what can be done and what is not feasible. Stakeholders, and by and large the public get to learn of the platform's community standards, and that there are buttons for reporting hate, misinformation and other social media bad behaviour. Most importantly, ways need to be developed in

[28] Ibid.

[29] Digital Action. 2022. Roundtable briefing: 2024: Global Year of Democracy.

[30] David Silverberg. 2023. Could AI swamp social media with fake accounts? https://www.bbc.com/news/business-64464140.

[31] Goldstein, Josh A et. al. 2023. Generative Language Models and Automated Influence Operations:Emerging Threats and Potential Mitigations. https://www.arxiv-vanity.com/papers/2301.04246/.

which stakeholders can contribute ideas to some of the challenges raised in a timely fashion. Platforms can receive a larger range of opinions from the user community by establishing collaborative, deliberative spaces. These can be through townhalls, and quarterly engagements akin to the convenings by KICTANet in 2022 in which platforms got to hear views from civil society actors.

Policy Recommendations

This paper makes the following recommendations to several stakeholders as follows:

Platforms

- In the context of the global election cycle in 2024 and considering that platforms have become sources of exchange of election information and disinformation, there is a need for preparation before the election years and periods of different countries. They need to put in place mechanisms for the challenges provided by the digital dangers on their platforms early enough and not necessarily during the election year. The mechanisms would include for example policies on how to deal with hate expressed in local languages, what to do with verified accounts of sitting leaders in the country if they are used for disinformation or hate speech, etc.
- Platforms must be deliberate in engaging with key stakeholders to inform them of what options are available for reporting electoral offenses and encourage them to give their views in order to strengthen the platforms. The engagements need to be regular and not confined to when elections are a few months away as there might not be adequate time to respond to new ways of doing things, or effecting changes recommended.
- Platforms will need to be intentional and create more awareness to let citizens know that they too have an option to participate, contribute and have their views taken on board. Communication to stakeholders should be simplified in such forms as animations, short videos, and in a way that inspires confidence in ordinary citizens (example here). In addition, there must be a way to demonstrate that citizens' opinions have been taken on board through clear feedback mechanisms, and not just taken through a participatory process as a mere ticking of boxes.
- Social media corporations, through their content moderation advisory groups, will need to deliberately consider users as citizens and collaborators in determining the direction in which the platforms determine content moderation models. The advisory groups' decisions need to have tangible effects, as currently, there is very little information on what these advisory or councils are doing or have been able to achieve.
- It is important for platforms to to be in tune with public values in order to respond accordingly.

Civil Society

- Future elections could see an even greater flood of false information. Some of these platforms such as Meta have an Eastern Africa office based in Nairobi, while others have no local presence, with their Africa policy teams based in Western Countries. Civil society therefore must be deliberate and strategic in engaging these platforms and point out local challenges. Further, they will need to make suggestions on safeguards, especially on the need for platforms to be accountable. In addition, they can also provide proposals on how context specific hate and disinformation on platforms can be minimized by respecting the local values that embrace safety and confidence of users wherever they are.

Donors

- There is a need for international development partners, donors and funders to avail resources to Global South Civil Society organisations in an effort to support their work in campaigns of engagement and calling out platforms on their lack of commitment to having safeguards and measures to combat disinformation and electoral lies on their platforms. Civil society organisations can push for the need to have policies in particular as they relate to the Global South when it comes to electioneering time. But this can only be achieved if resources are available to engage in meaningful and tangible work before, during and after a general election.

A Proposal on Voluntary Social Media Councils

Kuda Hove
INDEPENDENT RESEARCHER, SOUTH AFRICA

Introduction

Social media platforms have changed the way people connect with each other. Platforms such as *Facebook, Twitter, Reddit,* and *LinkedIn* have billions of users across the world including Zimbabwe. One thing that most mainstream social media platforms have in common is that they are developed and owned by entities based in the United States of America, Europe, or China in the case of TikTok. A platform's country of origin influences that platform's governance structure as well as the platform's policies. This is problematic given that these platforms are global, operating across a range of varied cultural and political environments and/or contexts.

Content moderation decisions often show how challenging it is to come up with one central content moderation policy which applies to a global range of users. For example, Facebook and Instagram initially absolutely barred the sharing of images and videos of topless women. This made sense from a Western perspective but in places like the Zulu nation in South Africa or in eSwatini, it is normal for girls and young women to take part in public cultural events topless. Sharing images of such cultural events was not possible as it violated the platform's community guidelines. In this instance, Facebook's community guidelines restrict Zulus and the Swati from exercising their full cultural expression. However, with time these community guidelines have evolved to include exceptions in which nudity is allowed under specific circumstances.

This is just one example of many where platform governance falls short of giving users, especially those outside of the US and Europe the ability to freely express themselves in a legitimate manner. Related to this, the global nature of platforms means that they are not easy to regulate or hold accountable when they contribute to the violation of fundamental rights in a jurisdiction such as Zimbabwe. This brief article proposes a form of social media platform governance that is suitable to protect the rights of Zimbabwe based users without unjustifiably restricting the enjoyment of their fundamental rights.

Can we learn from regulation of print media?

In Zimbabwe and other economically similar African countries, social media and instant messaging platforms have become sources of news and information seen by some as being of similar value as traditional print media sources. In a country with high inflation rates and low disposable income, most Zimbabweans cannot afford to buy print media in the form of magazines and newspapers. WhatsApp and TikTok based news sources and Facebook based newsgroups have mushroomed to inform audiences which otherwise would have been informed by traditional print media.

These online based platforms unfortunately can also be used to spread misinformation, manipulate voters during the run-up to elections or a key referendum, and even to fan genocide. This is why there is a need to moderate the content shared over social media platforms and also to put in place some form of regulatory frameworks to safeguard service users in a given jurisdiction.

To mitigate some of the harms which may be caused by the use of social media platforms, there has been a call for the introduction of effective oversight measures to ensure that content shared over those platforms does not amount to misinformation, hate speech or other forms of speech which may incite or

lead to the violation of certain fundamental rights. Social media platforms have, for example, come up with content moderation policies but one main shortfall and criticism of content moderation policies is that they are not flexible enough to apply fairly across the different cultural, political, and racial contexts within which one social media platform, such as Facebook operates. Some social media companies have also used their resources to create what may be referred to as social media councils. But these social media councils face the same problem as the platforms and their content moderation policies – they are not flexible enough to adequately apply to all global contexts.

It is my submission that similarities between print media and social media platforms mean that some regulatory print models for the print media can be modified and applied to social media platforms. An interesting prospect is the voluntary media council method as seen in the operations of the Voluntary Media Council of Zimbabwe (VMCZ).

The VMCZ is an independent, non-profit, self-regulatory body that promotes ethical and professional standards in Zimbabwe's media industry. The VMCZ was established by a group of media stakeholders who recognized the need for a self-regulatory mechanism to improve the quality of journalism, promote media freedom and more importantly, promote media accountability in the country. The VMCZ aims to promote a culture of ethical and professional journalism in Zimbabwe by providing a platform for media professionals and media consumers to engage in constructive dialogue on media ethics and best practices. Some of the country's largest independent media groups have volunteered to be bound by the VMCZ's Code of Conduct.

The VMCZ also seeks to promote media accountability through its complaints mechanism which members of the public can use to lodge complaints against media organisations or journalists who breach the VMCZ Code of Conduct they volunteered to be bound by. The VMCZ is not a regulatory authority and does not have the power to sanction or punish media organisations or journalists who violate ethical or professional standards. But the organisation does facilitate dialogue between the aggrieved party and the media organisation accused of breaching the VMCZ Code of Conduct.

VMCZ has managed to get some media journalists and media houses to print apologies and retract stories where they got the facts wrong or where the reporting breaches VMCZ's Code of Conduct. At a national governance level, VMCZ has been useful in lobbying the government for an independent Zimbabwe Media Commission.

What would a voluntary social media council look like?

Social media councils are often set up to perform an oversight role to evaluate the fairness of content moderation decisions taken by the social media platform they relate to. Statistics from Facebook's Oversight Board show that only 5 of the 46 complaints finalised between December 2020 and March 2023, were from Sub-Saharan Africa – specifically from Ethiopia, Sudan, Nigeria, and South Africa. This shows that a number of violations which take place on social media platforms go unreported if those violations take place in jurisdictions like Zimbabwe.

This could indicate that users are not aware of appeal processes which in turn may reflect a lack of awareness about the existence of social media councils like the Oversight Board. Setting up social media councils at national level could be useful in informing users about complaints and appeals procedures available to them. Additionally, establishing national level social media councils ensures that any local contexts are considered when investigating and handling a complaint.

The biggest challenge to the establishment of national level social media councils is the imbalance of power between most of the social media platforms and small economic markets like Zimbabwe, a low-income country. Social media platforms do not prioritise lower income countries because of the comparatively low return on investment they offer. High internet access costs in Zimbabwe mean that the number of users is comparatively lower than in bigger economies where people have more disposable income and enjoy higher levels of access to the internet.

Zimbabwe and other similar countries have no means to establish jurisdiction over foreign based social media platforms with financial reserves that are several times bigger than Zimbabwe's average national budget and GDP. As such, national social media councils are likely to be voluntary in nature with social media platforms choosing or volunteering to participate in these national media councils. Participation does not require that social media platforms establish a national office in each country, their participation can be virtual. In this way, content moderation teams can work with different national social media councils in each country to decide whether planned content moderation decisions are correct under the given circumstances.

Founding stakeholders for each national social media council stakeholders could be a mix of civil society actors, academics, users and representatives of special interest groups and minorities. The key is that these various stakeholders understand and have expertise in matters relating to free expression, access to information and related rights. It may also be useful to have public entities drawn from organisations such as National Data Protection Authorities, Free Speech Ombudsman and Constitutional Commissions tasked with promoting the enjoyment of the right to free expression and the related information rights.

A user driven regulatory approach is possible, as illustrated to a certain extent by Reddit's current content moderation process. Reddit's Content Policy is made up of eight rules which are universally applicable and stated in simple to understand language. These rules are enforced across each Reddit community not by a team of content moderators sitting in a regional Reddit office, but each community is moderated by its own community members who have been granted moderator status. These community level moderators have to abide by the terms of Reddit's Moderator Code of Conduct. This code of conduct provides oversight mechanisms for community moderators' activities.

How sustainable is this model?

The operational costs of these national social media councils would be minimal especially in instances when they are held virtually. Where possible, social media platforms could also support the maintenance of national social media councils since their work contributes to a safer user environment for each platform's users. Governments can also be instrumental in promoting the work that such national social media councils do. Research labs and civil society organisations can in a similar way support national social media council initiatives as a way of promoting their mission to protect free expression and information rights.

In instances when social media platforms do not voluntarily join national social media councils, these councils can play an advisory and referral role to global media councils such as Facebook's Oversight Board. In this way a Zimbabwean social media council would help users report violations of their privacy and refer those violations to the Oversight Board for redress. Additionally, national voluntary social media councils can also come together at regional levels to increase their chances of attracting validation/ the attention of social media platforms.

In conclusion, national social media councils, if executed well, can provide a legitimately independent, participatory, and transparent way to ensure user participation in content moderation and the regulation of social media platforms in each country. These social media councils would also be helpful in promoting awareness of safety measures and complaints processes across the various social media platforms.

Building and strengthening rights-based social media platform governance in Africa through national human rights institutions

Tomiwa Ilori
CENTRE FOR HUMAN RIGHTS, UNIVERSITY OF PRETORIA, PRETORIA, SOUTH AFRICA

Introduction

Most social media platform governance approaches in African countries are fraught with lack of trust and legitimacy. This lack is often due to the enormous rule-making and decision-making powers wielded by both African governments and social media platforms with respect to what stays on platforms and what does not. For example, the laws, policies and processes made by African governments to regulate online harms are often at variance with international human rights standards while social media platforms are distant from the contextual realities required to regulate online harms in African countries.

As a result of these, online rights are often violated at will while online harms continue to grow at an exponential rate. These violations and the growing threat of online harms have therefore made it necessary to rethink our idea of 'stakeholders' involved in the regulation of online harms beyond traditional government actors, social media platforms, and civil society in African countries.

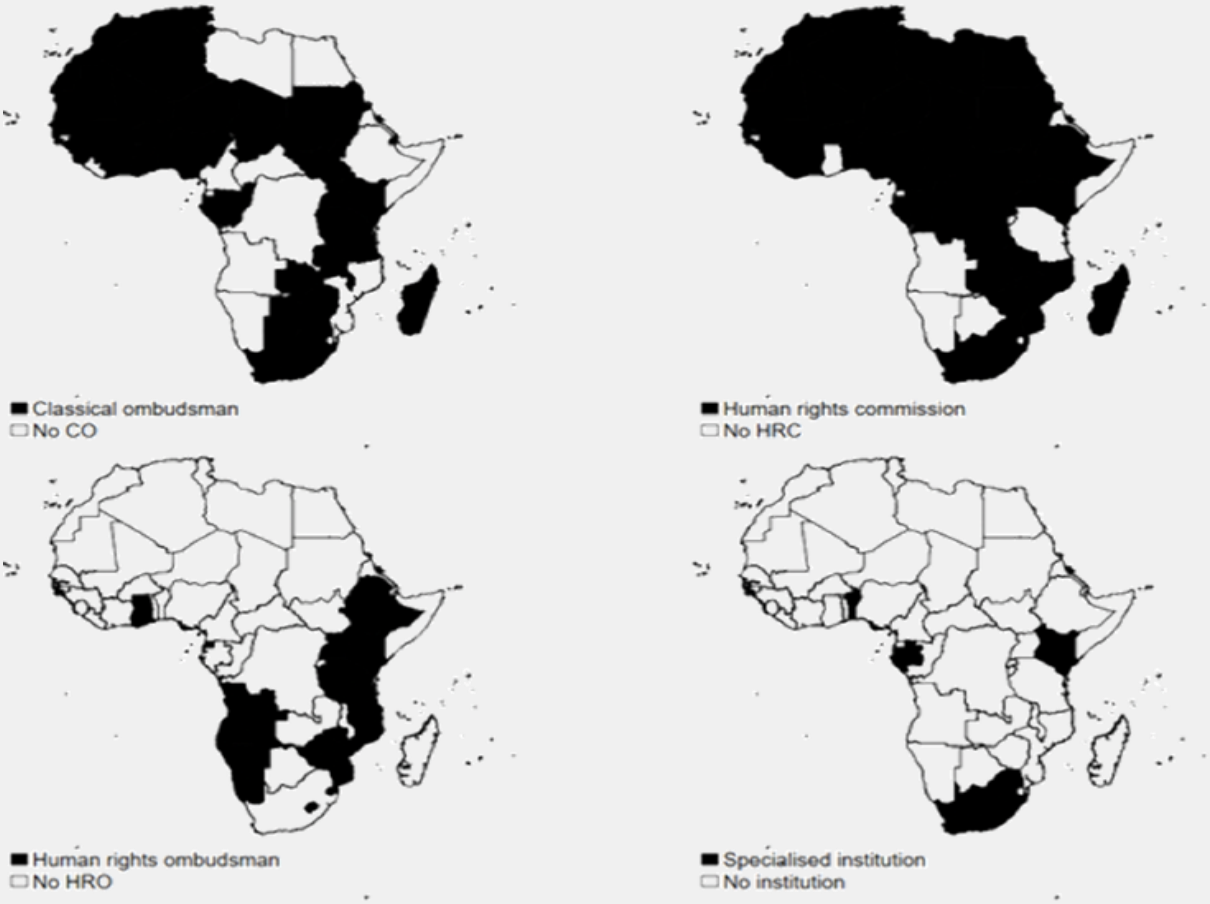
Missing from most ideas about social media platform governance in African countries, especially as it concerns protecting online rights and preventing online harms, is the legal mandate of national human rights institutions (NHRIs). Therefore, this contribution examines the roles African NHRIs can play to build and strengthen rights-based social media platform governance in African countries.

A brief overview of African NHRIs

NHRIs are unique statutory domestic institutions with a constitutional and/or legal mandate to protect and promote human rights in national contexts. According to Sundström, there are four types of NHRIs in African countries. They are:

(1) classical ombudsmen (COs), with a mandate focused on maladministration, (2) human rights commissions (HRCs), with a broad human rights mandate, (3) human rights ombudsmen (HROs), with a mandate of both maladministration and human rights, and (4) specialised institutions, such as children's ombudsmen, with a narrower mandate.

Types of African NHRIs



Source: Sundström, 2022

The first African NHRI was a classical ombudsman established in Tanzania in 1966. Currently, there are at least 47 NHRIs in African countries. Thirty-three out of the 54 African NHRIs have more than one type of NHRI. Most African NHRIs have both protective and promotional mandates and these include but are not limited to monitoring and reporting the human rights situation; providing advice to governments and others; delivering human rights education programmes; cooperating at the national level with key partners; and engaging with the international human rights system.

With respect to the last (but not the least) mandate, NHRIs have a key responsibility to ensure that key international human rights instruments are considered and implemented in local contexts. One of such key instruments is the United Nations Guiding Principles for Business and Human Rights (UNGPs) which has been further applied to states' obligations in ensuring rights-based social media platform governance in their various contexts.

Directed to social media companies, the human rights principles to be applied include ensuring human rights by default through terms of service, clear and determinable moderation rules, bases and extent of content moderation decisions, non-discrimination of protected characteristics, prevention and mitigation, transparency, and many others. Given NHRIs' mandates, they are well-positioned to monitor the implementation of these principles in domestic contexts by providing an anchorage for rights-based policies. One of such policies could be a soft law instrument such as a charter that sets out roles and

responsibilities of NHRIs and other actors with respect to the above-mentioned principles. However, despite having such a mandate and potential, NHRIs are faced with several challenges when it comes to social media platform governance in African countries.

Challenges facing NHRIs and social media platform governance in Africa

Social media platform governance in African countries is often seen from the vantage point of preventing online harms and protecting human rights. However, social media platform governance also includes economic and legal dimensions such as taxation, labour, competition and many others. But, so far as it concerns social media platform governance in African countries, human rights protection should rightly take centre stage for two major reasons. One, as critical stakeholders, many African governments have egregious internet freedom records, and two, power imbalances exist between states and social media platforms on the one hand and online publics on the other.

These reasons require us to think creatively about not just providing norms for social media platform governance but also think through how these provisions can be contextually relevant and effective domestically in African countries. In this regard, NHRIs are not only strategic domestic actors with respect to their main mandate to ensure national protection and promotion of human rights, they also serve as 'bridges' for mainstreaming international human rights standards such as those on social media platform governance into national contexts. Therefore, when it comes to the aspect of social media platform governance that involves human rights protection or mechanisms, including charters and social media councils, NHRIs need to play crucial roles.

However, African NHRIs face a number of challenges that could make playing these roles difficult. For example, as of November 2022, all the 28 African NHRIs rated based on the Paris Principles (Principles Relating to the Status of National Human Rights Institutions), which sets the minimum credible and operational standards that NHRIs must meet, received an 'A' status rating. This meant they were fully compliant with the Principles. However, out of this number, only four NHRIs including Ethiopia's Human Rights Commission, South Africa's Human Rights Commission (SAHRC), Nigeria's National Human Rights Commission (NHRC), and Kenya's National Commission on Human Rights (KNCHR) touch on the relationship between NHRIs and online rights protection as an objective in their annual strategic plans.

At the regional level, the Network of African National Human Rights Institutions' (NANHRI) Strategic Plan (2021-2025) highlights 'emerging and evolving issues' as some of the key strategic challenges facing NANHRI but does not expressly mention online rights protection. These challenges point to NHRI's lack of readiness to provide the capacity necessary for social media platform governance in African countries. In addition to this, NHRIs also have other capacity challenges related to financing, leadership, composition, human rights literacy and others.

Despite the challenges, it is not all gloomy. While, in non-African NHRIs have embraced their mandates in the facilitation of human rights protection in the context of data-driven businesses and recommendations for an Artificial Intelligence Safety Commissioner, African NHRIs such as the SAHRC and the NHRC are also seeking to exercise their mandate in the area. For example, SHRC is due to release a social media charter (SMCh) in South Africa, while NHRC is the main implementation actor in a proposed law on digital rights and freedoms in Nigeria. However, these taken together still point to a weak relationship between social media platform governance and NHRIs in Africa, as, in the larger context, it is almost non-existent, and what this points to, is the need to build and strengthen such a relationship.

Building and strengthening a rights-based social media platform governance through African NHRIs

The European version of NANHRI, the European Network of Human Rights Institutions (ENNHRI), has highlighted the debates in online content regulation and identified four main ways through which NHRIs can protect online rights and freedoms including providing advice, advocacy, implementing the Marrakech Declaration into the expanding civic space, and promoting and protecting human rights defenders, and monitoring and reporting in online and offline civic spaces. This suggests that not only should NHRIs build their domestic capacity, they should also work together as a strong unit at the regional level.

At the domestic level, efforts by NHRIs in South Africa and Nigeria point to norm-setting and decision-making with respect to social media regulation and broader digital rights issues. While such efforts are crucial, in addition to them, NHRIs need to also get actively involved in playing an important role to ensure a rights-based approach to social media platform governance in African countries. This way, NHRIs lead on building and strengthening the capacity of duty-bearers to meet their obligations while also encouraging rights holders to claim their rights. Some of the ways through which such a role can be played include strategic collaboration, policy advocacy, capacity building and radical participation.

Strategic collaboration: Granted that social media platform governance is a complex subject involving diverse actors and interests, such a complex subject can be brought apart and pieced together especially at the domestic level by NHRIs and other major stakeholders. This is because NHRIs are strategic actors with respect to domestic implementation of human rights and this strategic positioning also places them amidst other critical actors such as governments, regional institutions, international organisations, academia, civil society and others.

Such collaborations should provide a multi-stakeholder platform for addressing some of the issues addressed above. For example, NHRIs, in partnership with these stakeholders may publish and disseminate a guide for NHRIs on the regulation of digital technologies in Africa. The purpose of the guide would be to provide for the responsibilities of NHRIs in the regulation of digital technologies and also provide specifically for the role of NHRIs in social media platform governance in African countries.

Policy advocacy: NHRIs need to include policy advocacy for digital rights protection, including those for social media charters in their strategy documents. Such advocacy may include research, workshops, seminars with strategic stakeholders referred to above on how to embark on a digital rights legal reform in African countries. This would provide for adequate planning, execution and measurement of specific activities towards such regulation. Another means of carrying out such advocacy is through legal advocacy. Such advocacy will involve providing advice on the human rights impacts of proposed digital policy laws, and where necessary, support public interest litigation.

In addition to this, new social media regulatory mechanisms such as social media councils (SMCs) are now being proposed by international actors. SMCs are independent, multi-stakeholder ombudsman mechanisms modelled after press councils that enable industry-wide complaint mechanisms and promote remedies from violations. Seeing the attempt by the SAHRC to create a SMCh which is usually the basis for establishing SMCs and based on Sundström's typology of NHRIs, SMCs can be established as classical ombudsperson, human rights commissions, human rights ombudsperson or specialised institutions. However, any preference out of these types must be thoroughly critiqued to justify its adoption.

Capacity building: Capacity building in this context refers to equipping NHRIs with the ability to perform their roles in social media platform governance in African countries. This includes but is not limited to building and strengthening NHRIs' educational and financial capacity. With respect to educational capacity building, there is a need for more knowledge exchange among NHRIs and other stakeholders about the opportunities and challenges with respect to social media regulation in their countries.

Such knowledge exchange can be carried out through workshops, seminars and fora. With respect to financial capacity, African NHRIs should expand their strategic partnerships in order to source for more funds to address their financial constraints. This is because one of the means of guaranteeing the independence of public institutions such as NHRIs is their financial sustainability and independence. However, in order to safeguard this independence, NHRIs should develop and publish an ethical funding policy.

Radical participation: One of the biggest challenges facing democratic social media regulation across the world today is how to make it radically participatory. Such radical participation involves giving everyone opportunities to influence how social media platforms are governed. This radical participation is also becoming more necessary given the public power wielded by social media companies.

This is a mammoth task not only because of the sheer legal and administrative measures required to facilitate such participation but also because providing such radicality requires grassroots participation. While NHRIs do not have answers to all the problems of social media regulation in African countries, they are strategically placed to implement these measures and ensure such participation especially at the domestic level. One of the measures that could drive such participation are SMCs. However, whether they will ensure this kind of radical participation in African countries is still left to be seen.

Conclusion

In order to start the serious conversations about social media platform governance in Africa, we need to work with what we have. And with this in mind, rights-based content regulation cannot keep playing catch up with ever-evolving social media platforms. Therefore, we will need to devise creative means for domestic participation in social media platform governance – and NHRIs provide such means. Prior to the 2000s, data protection developments in Africa were limited. Building on constitutional provisions on the right to privacy as what we have, today, not only do 35 African countries have data protection laws, at least 18 of them now have unique institutional mechanisms -- data protection authorities. This is still very far from what a strong regional data protection landscape should look like, however, they are positive developments. With South Africa's SAHRC set to release a Social Media Charter, African countries are on another journey of what is possible with what they have -- NHRIs -- and it is high time they took a lead as strategic stakeholders in building and strengthening rights-based social media platform governance in Africa. This includes NRHIs building and strengthening strategic collaboration, policy advocacy, capacity building and radical participation.

Unpacking the Tensions: A Comparative Analysis of DNS Abuse Mitigation and its Impact on Business and Human Rights

Ephraim Percy Kenyanito
QUEEN MARY UNIVERSITY OF LONDON, LONDON, UNITED KINGDOM

Introduction

In today's digital age, the internet has become an integral part of our daily lives. With the ever-increasing use of the internet, it has become a powerful tool for communication and expression of thoughts and ideas. However, the unrestricted use of the internet has also brought forth numerous challenges, including DNS abuse, intermediary liability, and censorship, which affect the fundamental right of freedom of expression.

DNS abuse refers to the malicious use of the Domain Name System (DNS), which is responsible for translating human-readable domain names into IP addresses. DNS abuse includes activities such as phishing, malware distribution, spamming, and other malicious activities that exploit the DNS system. We note that DNS Abuse cannot be solved without imposing Intermediary Liability on DNS operators. Intermediary liability, on the other hand, refers to the legal responsibility of intermediaries, such as internet service providers (ISPs), for the content posted on their platforms. Finally, censorship involves the control of information flow on the internet, with the aim of suppressing or controlling certain types of content.

In three previous blogs, this author has highlighted the risks of regulating DNS Abuse without clear definitions and without respecting human rights.[1]

In this article, we will explore the complex issues surrounding the use of the internet, including DNS abuse, intermediary liability, and censorship, with a focus on the institutional, legal, and contractual frameworks in place to mitigate these challenges. We will also explore the impact of these challenges on the fundamental human right of freedom of expression.

Institutional Framework

The internet is a decentralized network of networks that operate under the stewardship of various organizations, including the Internet Corporation for Assigned Names and Numbers (ICANN). ICANN is responsible for managing the DNS system and ensuring that it operates smoothly and securely. The ecosystem surrounding ICANN is vast, comprising various stakeholders, including registries, registrars, and internet service providers (ISPs).

The role of these stakeholders is crucial in the implementation of DNS abuse mitigation measures and accruing liability. Registries are responsible for the management of top-level domains (TLDs), such as

[1] https://www.article19.org/resources/online-freedoms-safeguards-must-be-balanced-with-free-expression, https://www.article19.org/resources/icann-content-moderation-at-the-infrastructure-level-is-a-dangerous-move, https://www.article19.org/resources/europe-content-moderation-at-infrastructure-level-must-respect-human-rights.

.com, .org, and .net, while registrars act as intermediaries between registries and domain owners. ISPs provide access to the internet and are therefore responsible for ensuring that their networks are secure and free from DNS abuse.

To address DNS abuse, ICANN has implemented several policies and programs. One of these is the Contractual Compliance Program, which enforces the terms and conditions of ICANN's agreements with domain name registrars and registries. This program ensures that these entities comply with ICANN's rules and regulations related to DNS abuse.

ICANN also has a Global Domains Division (GDD) that works to protect the security and stability of the DNS. The GDD collaborates with other organizations, such as law enforcement agencies, to identify and mitigate DNS abuse. In addition, ICANN participates in industry groups and initiatives, such as the Anti-Phishing Working Group (APWG) and the Messaging, Malware, and Mobile Anti-Abuse Working Group (M3AAWG), to address DNS abuse.

In addition to ICANN, other organizations play a vital role in the management of the internet, including the Internet Engineering Task Force (IETF), the World Wide Web Consortium (W3C), and the Internet Governance Forum (IGF). The IETF develops and promotes internet standards, while the W3C develops web standards. The IETF has also taken steps to address DNS abuse. For example, the DNS Abuse Working Group, a subgroup of the IETF's Security Area, focuses on developing technical solutions to address DNS abuse. The working group has published several documents related to DNS abuse, including RFC 7706, which describes a method for identifying and mitigating DNS-based threats. The IGF, on the other hand, is an open forum for stakeholders to discuss internet governance issues.

Legal and Contractual Framework

Registry policies play a critical role in mitigating DNS abuse, intermediary liability, and censorship. The policies of select registries such as (.org[2]), (.ke)[3], (.cn)[4], (.eu)[5], (.br)[6], (.mw)[7], and (.tz)[8] have been analyzed against international human rights law and best practice.

The .org registry has implemented a domain abuse activity reporting system, which aligns with best practice for mitigating DNS abuse. The .ke registry is in the process of developing a draft policy on the prevention and mitigation of DNS abuse, which the author intends to analyse in future publications for its consistency with international human rights law. However we can note that previous actions at the registry indicated grounds for concern.[9]

---

[2] https://thenew.org/org-people/about-pir/policies/anti-abuse-policy/#:~:text=Technical%20Abuses%20of%20the%20DNS&text=This%20Policy%20prohibits%20the%20following,access%20to%20private%20computer%20systems.

3 https://kenic.or.ke/demystifying-domain-name-system-dns-abuse/ and KENIC Registrar Accreditation Agreement, https://kenic.or.ke/wp-content/uploads/2020/01/KENICS-REGISTRAR-AGREEMENT-1.pdf.

4 People's Republic of China, 2017, Article 3, https://www.cnnic.com.cn/PublicS/fwzxxgzcfg/201710/t20171026_69608.html.

5 Regulation (EC) No 733/2002 of the European Parliament and of the Council (3) and by Commission Regulation (EC) No 874/2004 (4).

6 Resolution CGI.br/RES/2008/008/P.

7 mw ccTLD Domain Registration Policy Version 1.2a of 23 July 2015.

8 Electronic and Postal Communications (Domain Name Management) Regulations, 2020.

9 https://www.article19.org/resources/icann-content-moderation-at-the-infrastructure-level-is-a-dangerous-move/.

The policies of the (.cn) registry have been criticized[10] for being inconsistent with international human rights law, particularly regarding freedom of expression. This criticism stems from the fact that through DNS policies, China is censoring access to information for citizens and residents within Chinese borders. The (.eu) registry has implemented a content removal policy, which has been criticized by this author before for potentially violating the right to freedom of expression and access to information.[11]

The policies of the (.br) registry have been praised for being consistent with international human rights law and promoting transparency and accountability in the management of DNS abuse.[12] The (.mw) registry has developed a domain name dispute resolution policy, which attempts to align with international best practice. We can however note that beyond analysing the policies themselves for these registries, there is lack of clear updated information on implementation of the policies and even in instances where an analysis of the policy notes that the policy is compliant with international best practices, there is still a chance that the implementation might be marred with loopholes to allow for violation of human rights policies.

The (.tz) registry has implemented a policy on domain name suspension and deletion, which has been criticized for lacking transparency and potentially violating the right to due process.[13]

Overall, registry policies have a significant impact on DNS abuse, intermediary liability, and censorship. It is important for registries to ensure that their policies are consistent with international human rights law and best practice to ensure the protection of human rights while mitigating DNS abuse.

Conclusion and Recommendations

In this article, we have examined the intersection of DNS abuse, intermediary liability, and human rights, with a particular focus on the right to freedom of expression. We have analyzed the institutional and legal framework surrounding DNS abuse mitigation measures, as well as the policies of select registries, against international human rights law and best practice.

Our analysis has highlighted that DNS abuse poses a significant threat to human rights, particularly the right to freedom of expression. While DNS abuse mitigation measures are necessary, it is crucial to ensure that they are consistent with international human rights law and do not result in censorship or violate other human rights.

We have identified key stakeholders within the ICANN and Internet Governance ecosystem who are involved in the implementation of DNS abuse mitigation measures and accruing liability. These stakeholders have a responsibility to ensure that DNS abuse mitigation measures are consistent with human rights and to promote transparency and accountability in their implementation.

Our analysis of select registry policies has highlighted the need for registries to ensure that their policies are consistent with international human rights law and best practice. While some registries have

10 https://link.springer.com/chapter/10.1007/978-3-031-28486-1_19 and https://digitalmedusa.org/wp-content/uploads/2021/08/Requiem-SSRN.pdf and https://www.tandfonline.com/doi/pdf/10.1080/23738871.2020.1805482.

11 https://www.article19.org/resources/europe-content-moderation-at-infrastructure-level-must-respect-human-rights/.

12 https://isoc.org.br/files/Study_on_the_Marco_Civil.pdf and https://link.springer.com/chapter/10.1007/978-3-030-63501-5_8.

13 https://www.article19.org/resources/icann-content-moderation-at-the-infrastructure-level-is-a-dangerous-move/.

implemented policies that align with human rights, others have been criticized for being inconsistent with human rights and potentially violating the right to freedom of expression.

Based on our analysis, we recommend the following:

1. Incorporate human rights considerations into DNS abuse mitigation measures: DNS abuse mitigation measures should be designed and implemented with a focus on protecting human rights, particularly the right to freedom of expression. This requires collaboration between stakeholders within the ICANN and Internet Governance ecosystem to ensure that DNS abuse mitigation measures are consistent with human rights.
2. Promote transparency and accountability in DNS abuse mitigation measures: It is essential to ensure that DNS abuse mitigation measures are transparent and accountable. This requires the development of clear policies and procedures, as well as regular monitoring and reporting on the implementation of these measures.
3. Conduct regular assessments of registry policies: Registries should conduct regular assessments of their policies to ensure that they are consistent with international human rights law and best practice. Where policies are found to be inconsistent with human rights, registries should take steps to revise and improve their policies.
4. Integrate human rights considerations into business decision-making: The business community has a significant role to play in promoting human rights in the management and handling of DNS abuse. Businesses should integrate human rights considerations into their decision-making processes and ensure that they are not contributing to human rights abuses through their actions or inactions.

In conclusion, DNS abuse poses a significant threat to human rights, particularly the right to freedom of expression. While DNS abuse mitigation measures are necessary, they must be consistent with international human rights law and not result in censorship or violate other human rights. We recommend the incorporation of human rights considerations into DNS abuse mitigation measures, the promotion of transparency and accountability in the implementation of these measures, regular assessments of registry policies, and the integration of human rights considerations into business decision-making. By taking these steps, we can mitigate DNS abuse while protecting human rights.

Centering Victims is Imperative for Effective Remediation in Platform Governance

Thobekile Matimbe
PARADIGM INITIATIVE, LAGOS, NIGERIA

Introduction

The civic space in some African countries is constantly shrinking due to crackdowns on journalists and civil society actors, causing concerns over their ability to express themselves freely.[1] Online gender-based violence is a serious pandemic for vulnerable groups such as women and children. Hate speech is unleashed on marginalised groups in Africa, and harm is perpetrated using social media platforms.[2] Increasing surveillance and arbitrary deployment of legislation to censor and halt the media results in censoring vulnerable groups from online platforms. Platform governance exists in this aggressive context where privacy and freedom of expression are constantly at risk.

A victim-centred approach is key in interventions that promote human rights, whether led by State or non-State actors. Drawing from the United Nations (UN), this approach puts the needs of victims and their safety first, includes a continuous and holistic approach to the delivery of services, and creates an enabling environment for the victims to be heard and supported.[3] Within the context of gross violations of human rights, the UN states that 'victims should be treated with humanity and respect for their dignity and human rights, and appropriate measures should be taken to ensure their safety, physical and psychological well-being and privacy, as well as those of their families.'[4] While quasi-judicial bodies do not have the same binding force as the courts, they can give important guidance in remedial action. In the context of platform governance, platform councils, such as the Oversight Board, may act as quasi-judicial bodies, giving persuasive guidance to the platform while lacking binding authority. The Oversight Board focuses on supporting free expression and other human rights, independently reviewing and making difficult decisions or giving advisory opinions about what content to leave up and take down.[5]

Social media platforms (SMPs) have policies and community standards that define which content is not acceptable on their platforms, and these purport to not only promote business expediency but safeguard the rights of users. Internet intermediaries are non-State actors and not parties to human rights treaties that bind States. Nevertheless, they do have considerable impact on human rights, and there are wide calls across society for affording them according responsibility to promote human rights, examples being transparency of processes addressing content moderation and effective remedial action for victims. Against this backdrop, once SMPs take on a corporate responsibility to protect human rights, platform councils such as the Oversight Board should also apply a victim-centred approach in discharging their function. Looking at the Oversight Board, this approach is lacking as their decisions

[1] Paradigm Initiative Londa 2021 Report https://paradigmhq.org/londa/ (accessed 19 February 2023).

[2] Africa renewal https://www.un.org/africarenewal/magazine/all-out-fight-against-hate-speech (accessed 28 March 2023).

[3] United Nations I have the right https://www.un.org/en/victims-rights-first (accessed 20 March 2023).

[4] United Nations https://www.ohchr.org/en/instruments-mechanisms/instruments/basic-principles-and-guidelines-right-remedy-and-reparation

[5] Oversight Board.

and advisory opinions are non-pecuniary and have not explored prescribing multistakeholder cooperation in providing psychosocial support for victims of harmful content. Without overstretching the role of the Oversight Board, substantively making their remediation more relevant to victims is necessary in effectively addressing user harm. This approach will also strengthen Meta's ability to consult the relevant stakeholders effectively and strengthen its community standards.

United Nations Guiding Principles

Bearing in mind fundamental rights are non-binding for non-State actors, a human rights lens is still instructive. The United Nations Guiding Principles on Business and Human Rights (UN Guiding Principles) set out the responsibilities of businesses towards human rights, and social media platforms fall within this category. Applicable human rights would then be freedom of expression as is guaranteed in article 9 and article 19(2) of the African Charter on Human and Peoples Rights and the International Covenant on Civil and Political Rights, respectively. However, the right can be limited in terms of a law, in pursuance of a legitimate aim and the means of limitation must be necessary and proportionate in accordance with international human rights standards. The Siracusa Principles and the African Commission on Human and Peoples' Rights Declaration of Principles on Freedom of Expression and Access to Information in Africa (the Declaration) are instructive here. SMPs should be transparent about how they address human rights and the remedial steps they take to address victims of harmful online content.

Remedial action is critical to the role of non-state actors as elaborated in the UN Guiding Principles.[6] The same stipulates the importance of an effective remedy for victims.[7] This approach calls for a victim-centred approach for SMPs and platform councils to take when addressing the human rights impacts arising from the actions of SMPs. Similarly, this approach should guide the adjudication process of platform councils in reviewing the conduct of non-state actors regarding human rights. In addition, the UN Guiding Principles stipulate that in providing remedies for business-related human rights abuses, mandates of existing non-judicial mechanisms can be expanded, citing national human rights institutions as critical.

Access to Effective Remedies

In Zimbabwe, women bear the brunt of online gender-based violence. Many are forced to remove themselves from online platforms. A report on online gender-based violence in Zimbabwe captures informant A's story of how she refrained from Facebook and WhatsApp for a year after her boyfriend shared her nude pictures online.[8] She did not get help from the police and was not aware whether 'Facebook' took down all her nude pictures taken without her consent. Many victims of the non-consensual sharing of intimate images face mental distress to the point where the breach of their privacy and security of person leads them to severe depression. Victims of online gender-based violence may need psychosocial support and effective remedies to address the harm caused. Social Media Platforms

[6] UN Guiding Principles on Human Rights and Business, Page 4https://www.ohchr.org/sites/default/files/Documents/Issues/Business/Intro_Guiding_PrinciplesBusinessHR.pdf (accessed on 19 February 2023).

[7] UN Guiding Principles on Business and Human Rights https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf (accessed on 19 February 2023).

[8] T Matimbe, Country Report on Zimbabwe Page 63 https://genderlinks.org.za/wp-content/uploads/2022/05/Understanding-Online-GBV-in-Southern-Africa-FINAL.pdf (accessed on 20 February 2023).

can address the prevalence of such cases by taking on advisory opinions that reflect prevalent harm to vulnerable groups and recommend effective remedial action beyond the appeal processes in the select privileged cases. In the case of the Oversight Board, issuing effective advisory opinions on how to meet the needs of victims of online gender-based violence and other harms online can go a long way in effective remedial action for victims. This approach has been deployed by the United Nations Refugee Agency in addressing the needs of victims of sexual misconduct where psychosocial support and risk assessments feature as a remedy for victims.[9]

The UN Guiding Principles aptly provide that situations may arise where a business enterprise requires active engagement in remediation, by itself or in cooperation with other actors.[10] The Due Diligence Project, in a submission to the report by the Office of the High Commissioner for Human Rights (OHCHR) on how to address the gender digital divide from a human rights perspective, highlights concerning online violence that internet intermediaries cannot be held liable for the initial violence, but obligations exist for both state and non-State actors.[11] As such, SMPs must consider the best form of redress for victims in their decisions. In platform governance, the remedies for harmful content that social media councils can order, are, for instance, taking down content for victims of online gender-based violence or reinstatement of content that does not violate community standards or policies.

Consideration of Cooperation Orders

In the Tigray Communication Affairs Bureau case (2022-006-FB-MR), the Oversight Board upheld Meta's decision to remove content threatening violence in the Ethiopian conflict.[12] However, the Oversight Board was concerned about the enforcement of Meta's content policies in times of armed conflict and ordered Meta to look at the feasibility of a sustained internal mechanism that 'provides the expertise, capacity and coordination required to review and respond to content effectively for the duration of a conflict'. While this decision addresses the non-recurrence of harm in conflict areas or attempts to resolve it, reliance on internal mechanisms alone does not suffice. An order for cooperation with other stakeholders in addressing violence online is a more wholesome approach to addressing harmful content.

Multistakeholderism is critical to limit the harms experienced through social media platforms. Consulting extensively with victim support groups and clinical psychologists is part of engaging different stakeholders in addressing victim needs, an approach that can be adopted by platform councils. The United Nations Office on Drugs and Crime (UNODC) highlights 'the right of victims to an adequate response to their needs', which is essentially a victim-centred approach that includes looking at support and assistance as important in remediation.[13] Platform councils must be empowered to give more effective remedies beyond acting as last-resort teams of content moderators where they handle appeals, as in the case of the Oversight Board. When content causes harm, an effective remedy for victims is

[9] UNHCR A Victim-Centred Approach https://www.unhcr.org/victim-care.html (accessed on 20 March 2023).

[10] UN Guiding Principles on Business and Human Rights https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf (accessed on 19 February 2023).

[11] Due Diligence Project Eliminating Online Violence Against Women and Engendering digital equality https://www.ohchr.org/sites/default/files/Documents/Issues/Women/WRGS/GenderDigital/DueDiligenceProject.pdf (accessed on 20 February 2023).

[12] Oversight Board https://www.oversightboard.com/news/592325135885870-oversight-board-upholds-meta-s-decision-in-tigray-communication-affairs-bureau-case-2022-006-fb-mr/ (accessed on 20 February 2023).

[13] UNODC Topic three - The right of victims to an adequate response to their needs https://www.unodc.org/e4j/zh/crime-prevention-criminal-justice/module-11/key-issues/3--the-right-of-victims-to-an-adequate-response-to-their-needs.html (accessed on 20 February 2023).

critical, and SMPs should be part of that process of addressing the harm as part of their corporate responsibility.

Social media councils can make decisions even if they are non-binding, but their function as quasi-judicial bodies must permit them to make key recommendations that go beyond community standards in addressing harmful content in line with human rights standards. In *Case 2021-011-FB-UA*, the Oversight Board found the decision by Facebook to remove a post discussing South African society under its Hate Speech Community Standard was appropriate as the post contained a slur deemed degrading, excluding and harmful to the people it targeted, within the South African context. This decision was victim centred as it considered a historical past of a group of people within South Africa and ruled that the slur in issue was against community standards. Beyond removing that content, the SMP decision would benefit from a recommendation of cooperation with the South African Human Rights Commission[14] and organisations promoting transitional justice to raise awareness of the retrogressive nature of hate speech shared on its platforms to foster a culture of human rights.

Leveraging Accountability Mechanisms

Platform councils like the Oversight Board lack the binding force to address the effects of harmful content effectively but can make advisory opinions and preside over appeals. Beyond this, they cannot address systemic online violence, and they need to cooperate with other accountability mechanisms within the different contexts and jurisdictions where harmful content originates. In the context of harm caused by political actors targeting vulnerable groups such as human rights defenders and journalists, working with national human rights institutions with powers to conduct investigations and address the State actors would help reduce the incidence of harmful online content. Cooperating with civil society actors will also ensure that the Oversight Board interacts well with the challenges faced by victims and orders custom-made remedial action that is victim-centric.

Conclusion and recommendations

There is a need for SMPs to adopt a post-content moderation collaborative remediation process that meets the needs of victims of harmful content. This is an innovative approach that SMPs can adopt to ensure they discharge their responsibilities in the human rights protection matrix. Platform councils can lead by prescribing this approach to foster better accountability from SMPs, demonstrating they care beyond financial expediency. Harmful content always comes at a cost, one way or the other. For the private sector, where they retain harmful content in the public domain for longer than necessary, their actions violate a victim-centred approach prolonging the harm caused. Remedial action should go beyond the removal of the harmful content to also address other forms deemed necessary by victims. In cases where platform councils get involved in the appeal of decisions made by SMPs, the victims, as a rule of practice, must be consulted about the nature of remedial action that meets their needs.

There should be consideration of psychosocial support for victims of grave online hate and online violence, for instance, supported by companies and other stakeholders to mitigate the adverse impacts of online content and promote peaceful societies. Companies must be proactive in their corporate social responsibility efforts to build a human rights-respecting environment. While companies are not directly

[14] A national human rights institution in South Africa.

liable for the pecuniary cost of online harm, the responsibility to promote human rights is not only for state actors but includes businesses. Platform councils, as non-judicial mechanisms to address content moderation, must prescribe evolving solutions to challenges in accordance with the nature of each case. Their decisions can also foster peace and security by promoting the responsible use of online platforms beyond community standards.

The following recommendations are critical to foster this approach:

- Decisions of platform councils must go beyond takedown of content and recommend cooperation of social media platforms with other accountability mechanisms in addressing the problem of harmful content.
- Platform council advisory opinions must proactively address prevalent cases of harmful content even where appeals are not lodged.
- Platform councils must consider cooperation with other non-judicial and accountability mechanisms to ensure a full appreciation of victims' needs in order to develop their decisions in human rights protection the way judicial bodies will advance their jurisprudence.
- Adequate and effective remedial action must consider promoting peace and security, psychosocial support for victims and other remedial action not limited to the obvious checklist of content moderation.

Platform Democracy, one size does not fit all: the case of GhanaWeb

Emmanuel Vitus
AW FREE FOUNDATION, ACCRA, GHANA

Competing interests among platform owners, users, and governments

In recent years it's increasingly becoming difficult for platforms around the world to identify and deal with harmful discourse. Such content often continues to gain traction online. People responsible for hateful speech online are constantly exploring new methods of bypassing the content moderation tools used by platforms to identify and control hateful speech online. This further complicates the challenges faced by platforms in addressing such content while also respecting freedom of expression. There are no easy answers. This paper looks at a particular example of how these challenges play out on GhanaWeb, Ghana's leading digital news and social media platform.

In Africa, a combination of factors makes content moderation particularly difficult, including colonial legacies, authoritarian governments, and the shrinking of civic space. Once renowned for their feisty media and defense of political liberties, Ghana's journalists and activists are under siege by the police. Online journalists and bloggers increasingly risk arrest and online harassment for their critical reporting.

Internet users in the West African nation can face criminal penalties for online content that is deemed to be false, which is broadly defined under existing law. In June 2021, members of Parliament proposed the Promotion of Proper Human Sexual Rights and Ghanaian Family Values Bill, 2021. If this legislation is passed, individuals who use internet services to produce or share materials advocating or supporting LGBT+ activities would face criminal penalties, including prison sentences of up to ten years. The proposed law was still being considered by Parliament.

In many African countries, due to questionable colonial laws that are now being transplanted into contemporary cybercrime legislation, platforms are now faced with complying with problematic laws that clearly violate free online speech. In Ghana, a new law passed in December 2020 which expands the government's legal authority for surveillance, despite existing data protection policies.

Much grassroots political activism in Africa relies on the use of U.S.-based social media platforms. However, these platforms also play host to state-backed manipulation efforts and can be subject to draconian shutdowns if the political dialogue goes awry for African governments. As a result, these companies get caught up in a tenuous position. To make matters even more difficult, there is an inescapable tension between the platforms' desire to apply global standards, mostly based on algorithms, to content moderation adjudication on the one hand and, on the other, to defer to local contexts when moderating content.

In the midst of this ruckus between platform owners, users and government, GhanaWeb, one of Africa's oldest websites, stands out in its approach to balancing the freedom of speech of its audience and the growing global trends to moderate online content perceived as potentially harmful.

Ghana's leading platform: GhanaWeb

GhanaWeb is Ghana's leading digital news and social media platform. It is an open platform launched in 1999 that operates under the laws of the Netherlands, a legal identity that allows Ghanaians to express themselves freely through opinion articles, multimedia content, and comments. The platform also hosts millions of user-generated posts published through the GhanaWeb Reporter, which is a digital media sharing platform integrated into the GhanaWeb App to give citizen journalists and content creators direct access to the means to publish a wide range of local content.

According to the Alexa website traffic statistics, GhanaWeb is very popular among Ghanaian diaspora in the United States, United Kingdom, Germany, Canada, Italy, South Africa, Netherlands, France, and many other countries.

GhanaWeb's history spans over two decades. It started in 1992 as a print magazine called GhanaHomePage serving the growing Ghanaian community in Amsterdam by collating news stories from Ghanaian newspapers. The GhanaWeb website was launched on January 1, 1999. In 2001, GhanaWeb was the first African website offering a mobile website which was initially supported by feature phones such as the iconic Nokia 3110. Through the Wireless Application Protocol (WAP), GhanaWeb went mobile six years before the first iPhone hit the market.

The portal grew steadily with many people helping to update the website with daily news and information from Ghana. Notable journalists, bloggers and many columnists have gained popularity through their work and opinions on GhanaWeb.

Out of the hundreds of active news and social media platforms in Ghana, GhanaWeb has by far the largest content moderation operation since the 2020 presidential election in the country. The platform's move followed the vote of the Cybersecurity Act, 2020 which provides broad authority for the Cybersecurity Authority (CA) to block or filter online content on receipt of a court order. The law also places penalties on service providers that fail to comply with a blocking or filtering order, which may include one to five years' imprisonment.

In addition, it is the platform that has come under the most scrutiny by the government for its content moderation decision-making practices, both human and automated. One of the reasons for this is that GhanaWeb is the largest news and social media platform in the West African country. It ranks third in national internet engagement after YouTube and Google.com since 2008.

GhanaWeb's content moderation policy

As a result, GhanaWeb's content moderation practices affect a significant amount of user expression. GhanaWeb utilizes both a centralized and hybrid approach to content moderation. However, 90% of the content moderation work is done by the community using a flagging system. The flagging mechanisms allow users to express their concerns by raising a red flag to report offensive or abusive content.

According to the CTO of Ghanaweb, the platform has a strict policy on comment moderation: 'Comments are moderated but not deleted on the platform.' This is in line with the total freedom of expression policy and values of GhanaWeb on comments and feedback from readers.

"*We believe that, at the very least, everyone has the right to be heard. It is not our job to shield the public from unpopular ideas. We provide the platform for as many Ghanaians as are willing to speak out to get their voices heard. Of course, in promoting freedom of expression we are alive to our responsibility to ensure that we do not promote hate speech, criminality, immoral conduct and any forms*

of expression that will jeopardize national security and cohesion. However, we remain as impartial and independent as possible from politics and remain a platform for the voiceless either through contents or comments", said the Editor-in-Chief of the platform.

In his words, GhanaWeb is an independent platform that owes allegiance to no political party, business entity or ethnic or religious grouping. *"Our first loyalty is to our audiences who turn to us every day in search of news and information to help them live their best lives and make good decisions. We don't take sides in any argument or debate. Our job is to ensure that all sides of every argument are laid bare before our audiences to help them make informed choices. We do not give any side more prominence in order to promote any particular agenda."*

Community to Combat Hateful Discourse

In an attempt to ensure that its internal policies can be enforced appropriately GhanaWeb tries to involve the community in a flagging system allowing the community of its over three million loyal visitors to do peer checking of comments and raise a red flag on comments that do not respect the community standard. This includes comments in local languages and icons. It's the first approach to comment moderation. However, once a comment reaches a certain threshold (number of flags), moderators reviewers who receive the same general training on the company's Community Standards and how to enforce them will evaluate the comments and make a decision. *"These guidelines are meant to help the editorial staff and moderators of GhanaWeb live up to our values. They are meant to guide our decisions as to what we keep or moderate and how we moderate them. They are guidelines, not doctrine or dogma. Every situation needs to be appreciated on its own facts and carefully considered on those facts and the discretion of the editorial team. In all situations, moderators are encouraged to have frank and open discussions among themselves with the aim of arriving at the best possible solutions in the interest of all concerned, the audience and the platform*," said the Editor-in-Chief.

Comments that get blocked can still be visible to those who choose to see them. The system places a caution wall on all comments that are flagged. According to owners of the platforms, commenters don't have to register to comment, they can't contest because they are anonymous. However, those who flag have to register to flag the comments. This makes the content moderation process transparent.

The language bit

The approach to language in Ghana is very important for GhanaWeb, because harmful content continue to gain traction online and those perpetuating it are constantly using new methods of getting around moderation tools, such as using local languages, symbols, memes and emojis to "disguise" the content so that the content moderation algorithm does not identify it as being problematic.

Ghana is a multilingual country in which about eighty languages are spoken. Of these, English, which was inherited from the colonial era, is the official language and lingua franca. Of the languages indigenous to Ghana, Akan is the most widely spoken in the south. Dagbani is most widely spoken in the north.

As part of its hybrid approach to content moderation, the GhanaWeb editorial team engages in several phases of technical and human review in order to identify, assess, and take action against content that potentially violates its Community Standards through using local languages.

Automated content moderation and other methods

In response to growing pressure from stakeholders such as government and the public to take down content that violates community standards quickly, GhanaWeb is now investing heavily in automated tools for content moderation. These include image recognition and matching tools to identify and remove objectionable content such as terror-related content and discriminatory comments directed at particular tribes or communities using language matching tools that seek to recognize and learn from patterns in text related to topics such as propaganda and harm. Depending on the level of complexity and the degree of additional judgment needed, the content may then be relayed to human moderators from the GhanaWeb editorial team.

As time passes, local platforms will continue exploring news methods to address the issue, however, policy-makers should resist the siren call of tighter regulation and illiberal measures. Instead, efforts should be made to mitigate the more damaging effects of social media in ways that take into account local information environments. This could include formulating more targeted measures to improve digital literacy and building trust in key institutions, such as the traditional media, which are often best able to act as arbiters of the truth locally.

The future of GhanaWeb

The case of GhanaWeb shows how local platforms in Africa are struggling to build a transparent and strong moderation system to enhance user expression/experience and also keep the platform safe and democratic. In order to align it's content moderation with the global trends and best practices, GhanaWeb must:

Balance moderation with censorship

Given the scale and reach of GhanaWeb, its content moderation policies need to account for the societal harms that can result from the mass distribution of hate speech and misinformation. The platform owners have a responsibility not to curtail speech too aggressively. Since hate speech and misinformation can be difficult to define, excessively restricting the reach of contentious political speech risks unduly limiting the freedom of expression on which democratic discourse depends.

Publish content guidelines and policies

GhanaWeb should clearly disclose what their content moderation policies are. Ideally, the policies would also be easy for users to understand and include either examples or clarifications of how ambiguous terms will be interpreted. Clear guidelines need to exist about what categories of content or comment will be restricted. Without public and transparent guidelines, content moderation decisions will appear ad hoc and undermine user trust.

Disclose moderation practices and appeals process

GhanaWeb owners should publicly and transparently disclose high-level details about their content-moderation practices, as well as their review process and publish clear guidelines for how to contest a moderation decision: If a comment or content has been banned, users have a right to know how to appeal that decision and whether the review process will involve an automated or manual review.

Disclose Algorithms

GhanaWeb owners should publicly and transparently disclose basic factors about what kinds of data the algorithm considers when flagging or removing a comment or comment on the platform.