ARTICLE 19

unesco
In partnership with UNESCO

Funded by the European Union

Content Moderation and Local Stakeholders in Kenya

June 2022

SOCIAL MEDIA 4PEACE

ARTICLE 19

T: +44 20 7324 2500
F: +44 20 7490 0566

E: info@article19.org
W: www.article19.org

Tw: @article19org
Fb: facebook.com/article19org

© ARTICLE 19, 2022

This publication was produced with the financial support of the **European Union** and **UNESCO**.
The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO or the European Union concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries.

The authors are responsible for the choice and the presentation of the facts contained in this publication and for the opinions expressed therein, which are not necessarily those of UNESCO or the European Union and do not commit the Organizations.

This work is provided under the Creative Commons Attribution-Non-Commercial-ShareAlike 3.0 licence. You are free to copy, distribute and display this work and to make derivative works, provided you:

1) give credit to ARTICLE 19;

2) do not use this work for commercial purposes;

3) distribute any works derived from this publication under a licence identical to this one.

To access the full legal text of this licence, please visit:
https://creativecommons.org/licenses/by-sa/3.0/legalcode

Acknowledgements

Many thanks to Victor Kapiyo, who has conducted this research and authored this report in close consultation with ARTICLE 19.

Victor is a lawyer practising in the Kenyan bar, in the firm of Lawmark Partners LLP, and a researcher on how emerging ICT public policy issues intersect with human rights in areas such as Internet freedom, privacy and cybersecurity. He is a Trustee of KICTANet, a researcher for the annual CIPESA State of Internet Freedom in Africa reports and the Vice-Chairperson of Youth Alive Kenya. He is also a member of the Law Society of Kenya. the Kenyan Section of the International Commission of Jurists (ICJ Kenya), and represents KICTANet in the Advisory Network of the Freedom Online Coalition (FOC-AN).

Many thanks to all the people who have shared their expertise and views and contributed to the research process.

Thanks also to our partners at UNESCO who have provided input and feedback at various stages of the development of this report.

Contents

Executive summary 5

Introduction 7

About the project 7

Methodology and structure of the report 9

Kenya at a glance 10

The state of content moderation in Kenya 15

Social media landscape in Kenya 15

Overview: Impact of content moderation on peace and stability 16

Issues with content moderation in Kenya 25

Impact of global content moderation practices at the local level 31

Interim conclusion 39

Analysis of stakeholders 41

Public sector actors 41

Civil society organisations 42

Social media companies 45

Academia 47

Media 48

Private actors 49

Religious actors 50

Conclusions 53

Recommendations 55

Annex A: Risk analysis 58

Annex B: Potential members of the coalition 63

Annex C: Interview sheet 67

Annex D: ICT status in Kenya 69

Bibliography 70

Endnotes 77

Executive summary

This study by ARTICLE 19 is conducted as part of the United Nations Educational, Scientific and Cultural Organization's (UNESCO) EU-funded project, **Social Media 4 Peace**, which aims to strengthen the resilience of civil society to potentially problematic content spread online, in particular hate speech inciting violence, while enhancing the promotion of peace through digital technologies, notably social media.

Kenya is a digital country with an Internet penetration of 93.7% and an increasing number of social media users (11 million to date). In this context, while remaining aware of the opportunities presented by social media platforms to freedom of expression and to serve as a platform for a peaceful and democratic society, the study documented issues with current practices of content moderation and their impact on peace and stability in Kenyan society. The spread of harmful content, in particular disinformation and misinformation; hate speech; online gender-based violence content; and malicious, coordinated, and inauthentic behaviour are key challenges, especially during electoral periods and, in particular, the upcoming August 2022 elections in Kenya. The study has identified flaws in content moderation practices which includes a lack of country-level data; algorithms that prioritise and amplify extreme, divisive, and polarising content; low public awareness and limited access to content rules in local languages; ineffective complaint mechanisms and remedies; marginalisation and exclusion of communities; lack of consideration for the various dimensions of the local context in content moderation practices; and inconsistent application enforcement of content rules.

The study mapped the capacity, knowledge, and needs of various stakeholders from government, civil society, private sector, and academia in relation to their work on content moderation.

The study also explored how sustainable and open engagement with local stakeholders could help social media companies to integrate a stronger understanding of the various dimensions of the local context into their content moderation systems, which would in turn improve the moderation of problematic content that negatively impacts the Kenyan society. The research submitted to local stakeholders the idea that a local multi-

stakeholder Coalition on Freedom of Expression and Content Moderation could play a positive role in that respect while ensuring that international human rights standards are duly taken into consideration in content moderation.

Most of the respondents welcomed the idea of a coalition that would work on the issues discussed and identified in the study. Such a coalition could be a useful platform to organise, engage, and co-create local, strategic solutions and responses to tackle the spread of problematic content on social media in Kenya. In addition, it could provide a useful avenue for engagement with social media companies to promote international human rights standards, transparency, and accountability in content moderation. The contributions from interviewees have allowed the elaboration of recommendations on how to facilitate the establishment of a local multi-stakeholder Coalition on Freedom of Expression and Content Moderation in Kenya.

Introduction

This publication has been produced as part of the United Nations Educational, Scientific and Cultural Organization's (UNESCO) project **Social Media 4 Peace** funded by the European Union (EU).

About the project

This report is part of the **Social Media 4 Peace** project that UNESCO is implementing in Bosnia and Herzegovina (BiH), Kenya, and Indonesia, with support of the EU. The overall objective of the project is to strengthen the resilience of civil society to potentially harmful content spread online, in particular hate speech and disinformation, while protecting freedom of expression and contributing to the promotion of peace through digital technologies, notably social media. ARTICLE 19's contribution to the project focuses on concerns raised by the current practices of content moderation on dominant social media platforms in the three target countries.

ARTICLE 19 considers that social media companies are, in principle, free to restrict content on the basis of freedom of contract, but that they should nonetheless respect human rights, including the rights to freedom of expression, privacy, and due process. While social media platforms have provided opportunities for expression, a number of serious concerns have come to light. The application of community standards has led to the silencing of minority voices. The efforts of tech companies to deal with problematic content are far from being evenly distributed: for instance, it has been shown that '87% of Facebook's spending on misinformation goes to English-language content, despite the fact that only 9% of its users are English speaking.' It has also been revealed that most resources and means in terms of content moderation are being allocated to a limited number of countries. Generally speaking, the transparency and dispute resolutions over content removals have so far been inadequate to enable sufficient scrutiny of social media platforms' actions and provide meaningful redress for their users. Finally, it is doubtful that a small number of dominant platforms should be allowed to hold so much power over what people are allowed to see without more direct public accountability.

This report specifically looks at the situation of local actors who, while they are impacted by the circulation of harmful content on social media or the moderation thereof, often find themselves unable to take effective action to improve their situation in that respect. They may feel frustrated by the inconsistencies of platforms' application of their own content rules; they may feel that global companies ignore their requests or misunderstand the current circumstances of the country or region. Some may lack understanding of content rules or of content moderation, but that is not the case of all local stakeholders.

The research then seeks to test, through the views of local stakeholders, the assumption that a local Coalition on Freedom of Expression and Content Moderation could play a role to fill the gap between the realities of local actors and companies that operate on a global scale. The idea for such a coalition is based on ARTICLE 19's work on the development of Social Media Councils, a multi-stakeholder mechanism for the oversight of content moderation on social media platforms. ARTICLE 19 suggested that Social Media Councils should be created at a national level (unless there was a risk that it would be easily captured by the government or other powerful interests) because this would ensure the involvement of local decision-makers who are well-informed of the local context and understand its cultural, linguistic, historical, political, and social nuances. While the development of a self-regulatory, multi-stakeholder body such as a Social Media Council is a long-term and complex endeavour, a local Coalition on Freedom of Expression and Content Moderation would be a lighter approach that could be supported within a shorter timeframe. Basing its work on international standards on freedom of expression and other fundamental rights, such a coalition could provide valuable input to inform content moderation practices, notably through its knowledge and understanding of the local languages and circumstances. As a critical mass of local stakeholders, it could engage into a sustainable dialogue with social media platforms and contribute to addressing flaws in content moderation and improving the protection of fundamental rights online. The coalition could provide training and support on freedom of expression and content moderation to local civil society actors that are impacted by content moderation. Finally, it could possibly pave the way to the creation of a Social Media Council in the country at a later stage. Through this research, at the initial stage of the **Social Media 4 Peace** project,

the idea of a local Coalition on Freedom of Expression and Content Moderation was submitted to local stakeholders, whose views have enabled the formulation of recommendations on how to approach the facilitation of a pilot coalition in the specific context of Kenya. In order to guarantee the effective ownership of the coalition by its members, the process facilitating its creation will necessarily include a validation exercise that ensures that potential members have the opportunity to discuss the findings of the research.

For the purposes of this report, we rely on the following definitions:

- **Content moderation** includes the different sets of measures and tools that social media platforms use to deal with illegal content and enforce their community standards against user-generated content on their service. This generally involves flagging by users, Trusted Flaggers or 'filters', removal, labelling, down-ranking or demonetisation of content, or disabling certain features.
- **Content curation** is how social media platforms use automated systems to rank, promote, or demote content in newsfeeds, usually based on their users' profiles. Content can also be promoted on platforms in exchange for payment. Platforms can also curate content by using interstitials to warn users against sensitive content or applying certain labels to highlight, for instance, whether the content comes from a trusted source.

Methodology and structure of the report

This report combines a policy and literature review conducted through desk research with qualitative interviews with 31 key informants comprising of policymakers, researchers, academics, bloggers, journalists, and human rights defenders (see Annex C for full list of interviewees).

For the interviews, an initial list of interlocutors was elaborated through desk research and in consultation with UNESCO with the aim of speaking with a broad and representative range of relevant stakeholders, including representatives from civil society, the private sector, public actors, and social media companies. However, in spite of the researcher's

best efforts, it was not possible within the timeframe of this research to arrange interviews with all the organisations identified originally.

The desk research allowed for the identification of issues linked to the circulation of problematic content on social media in Kenya. The identified content moderation issues were then discussed during the interviews, which aimed at understanding the experiences and challenges of Kenyan groups in dealing with platforms on content moderation issues. The idea of a local Coalition on Freedom of Expression and Content Moderation was also submitted for discussion with the interviewees, who provided their views on the overall idea of a coalition as well on the potential structures, members, roles, and dynamics of the coalition.

The Introduction highlights the diversity and complexity of the Kenyan society.

The second chapter describes the landscape of social media platforms and explores the dynamics related to the use of social media and the practices of content moderation related to Kenya. This chapter highlights the need for social media to integrate an informed understanding of the multiple dimensions of the local context when applying their global content rules.

The third chapter provides an analysis of the different stakeholder groups that deal with or are impacted by content moderation practices.

After the conclusions, the report puts forward recommendations based on interviews on how to facilitate the formation and operation of a civil society Coalition on Freedom of Expression and Content Moderation to reinforce an effective dialogue between social media and local civil society actors.

Kenya at a glance

Kenya is a culturally diverse country with over 70 distinct ethnic communities, speaking close to 80 different dialects and practising different cultural beliefs and traditions. English and Swahili are the official languages spoken in the country, with Swahili being spoken by the majority. According to the 2019 Census, the population stood at 47.6 million

with the largest ethnic community recorded being the Kikuyu (8.1 million), followed by the Luhya (6.8 million), Kalenjin (6.4 million), Luo (5.1 million), and the Kamba (4.7 million). The minority indigenous communities are the Dahalo (575), Gosha (685), El Molo (1,104), Konso (1,299), and Makonde (3,764). Furthermore, 85.5% of the population identify as Christians, while Muslims constitute 11% of the population.

Kenya's ranking in the 2021 Global Peace Index improved from 143 (score of 2.530) in 2011 to 116 (score of 2.254), ranking 27th in sub-Saharan Africa. Kenya is classified as a hybrid regime[1] and is ranked 95th globally and 13th in Africa in the 2020 Economist Intelligence Unit's Democracy Index, which reported a decline in the country's rankings from 5.11 in 2018 to 5.05 in 2020. The period following independence in 1963 has been marked by a history of domestic tensions and contestation associated with centralisation, high levels of corruption, and post-election violence. These issues fuelled the political crisis that ensued following the 2007 general election, which also resulted in widespread violence and the loss of 1,333 lives, the displacement of 650,000 people, and the destruction of property worth millions of US dollars. According to the Global Peace Index 2021, the economic cost of violence as a percentage of Kenya's gross domestic product in 2020 was 4%, equivalent to USD 9,013.1 million.

Kenya was ranked in the 102nd position in Reporters Without Borders' 2021 World Press Freedom Index, who also noted that respect for freedom of expression:

> *"depends a great deal on the political and economic environment, which was undermined by the coronavirus crisis in 2020. (…) Politicians continue to exercise a great deal of influence over both state and privately-owned media, which censor themselves to a significant degree, avoiding subjects that could cause annoyance or might jeopardise income sources. (…) Investigations into violence or abuses against journalists are still very uncommon and, as local NGOs point out, rarely lead to convictions. Journalists can pay dearly for covering opposition events or for portraying President Uhuru Kenyatta's party and its flaws in a negative light."*[2]

The 2007 political crisis was a watershed moment that led to the initiation of mediation efforts led by former UN Secretary-General Kofi Annan, resulting in an agreement for the

formation of a coalition government between then President Mwai Kibaki and opposition leader Raila Odinga and a raft of reforms. In August 2010, a progressive constitution which entrenched wide institutional reforms and embedded international human rights standards was adopted.[3] Further, in 2008, a Truth, Justice and Reconciliation Commission (TJRC) was established to document various systemic human rights violations and abuses by the colonial and the successive post-independence regimes in Kenya. The TJRC report, whose recommendations are largely unimplemented, detailed the violations and atrocities committed by these regimes from a historical perspective.[4]

In the run-up to the 2013 election, the challenges identified by TJRC together with demands for accountability for human rights violations arising from the 2007 post-election violence were evident in the campaigns. During this election, hate speech and propaganda in the forms identified in the 2007 election migrated from SMS and radio to email, online blogs, and social media platforms such as Facebook and Twitter. However, there was great caution, especially by mainstream media who primarily advanced what has been called a 'peaceocracy' narrative that focused coverage aligned to the government position on the need for a peaceful election, the acceptance of the election results and not protest, while carefully avoiding content that could potentially trigger conflict.

Social media platforms provided an avenue for the public to engage in discourse on election matters missing from mainstream media and to express discontent, criticism, and hold leaders accountable. President Uhuru Kenyatta and his Deputy, William Ruto, who in 2011 had been charged before the International Criminal Court for crimes against humanity, won the disputed poll. Raila Odinga, citing 'massive malpractice', challenged the results of the divisive election in court but the Supreme Court upheld the election stating it had been 'free, fair, transparent and credible'. In 2014 and 2016, the cases at the International Criminal Court against Kenyatta and Ruto, respectively, were terminated and charges withdrawn due to insufficient evidence.

Consequently, the unresolved historical tensions and political divisions, including those arising from the 2007 and 2013 elections, on where to strike the balance between the nexus of peace, conflict, consensus, debate, right to protest, democracy, justice, and free

and fair elections were replayed in the run-up to the August 2017 election. Likewise, the tensions between freedom of speech and offensive speech in a deeply divided context remained of concern. During the period, the National Cohesion and Integration Commission (NCIC)[5] monitored offline and online hate speech and issued warnings against inflammatory speech. The government's bid to control information saw officials attempt to obstruct critical journalists, bloggers, and civil society seeking accountability and speaking out against impunity with legal, administrative, and informal measures, including threats, intimidation, harassment, and online and phone surveillance. In some cases, physical assaults were also documented. The use of social media, including for propaganda, was more robust with Kenyatta's Jubilee Party deploying Cambridge Analytica to develop and run its online campaign, which the opposition described as 'worrying' and a 'criminal enterprise which clearly wanted to subvert the will of the people – through manipulation, through propaganda'.

Following the election, the country faced another political crisis as Raila Odinga challenged the election result at the Supreme Court of Kenya. The Court, in an unprecedented decision, nullified the result in what was a first by a court in Africa, plunging the country into uncharted waters. Kenyatta was subsequently elected president in October 2017, following a rerun of the Presidential election, which Odinga boycotted. The ensuing period was characterised by heightened political uncertainty and tension, which culminated in Odinga's swearing-in as 'the people's president' in January 2018.

In March 2018, the duo shocked the country with a political truce or 'handshake' and a political agreement dubbed the 'Building Bridges Initiative' aimed to unite Kenyans after the divisive 2017 elections that once again had left the country highly polarised. The controversial handshake cooled political temperatures but did not resolve feelings of alienation and marginalisation. The President appointed a Unity Advisory Taskforce which developed a report outlining nine priority areas of reform, which were consolidated into the Constitution of Kenya (Amendment) Bill, 2020. However, the adoption of the Bill was halted in May 2021 by the High Court, by the Court of Appeal in August 2021, and finally on 31 March 2022 by the Supreme Court for its unconstitutionality. The decisions altered the

prospects of prominent political figures and has recalibrated the country's political dynamics.

Against this background, it is likely that the general elections in 2022 will represent a period of tension in Kenya. Social media remains an indispensable tool in the political space especially during elections. A 2020 study found that social media networks, such as Facebook, WhatsApp, and Twitter, were used to spread fake news during the August and October 2017 elections with the objective of influencing voters, damaging the reputation of key institutions, fuelling political tensions and violence, and reaffirming existing prejudices that reflect social and political divisions.

Therefore, monitoring and combating the evolving nature and spread of hate speech, fake news, and online propaganda will remain key priorities for national regulators ahead of the August 2022 election, and more so the NCIC, which has already set up a unit to monitor and deter hate speech on social media platforms. Likewise, the need to address other types of problematic content online remains crucial, even as debates rage over the roles of social media companies, governments, and users over where to strike a balance on the appropriate level, scope, and extent of content moderation of social media content.

The state of content moderation in Kenya

Social media landscape in Kenya

Kenya's population grew from 46.4 million in 2018 to 49.8 million in 2021, at an annual rate of 2.3%. Likewise, mobile subscriptions increased from 45.6 million in June 2018 to 64.4 million in June 2021, while mobile (SIM) penetration increased from 97.8% to 132.2% (see Table 1 in Annex D). During the same period, Internet subscriptions increased from 41.1 million to 46.7 million; Internet penetration rose from 88.5% to 93.7%; broadband subscriptions increased from 20.5 million to 27.5 million; international Internet bandwidth increased from 3,278 Gbps to 10,218 Gbps; mobile connections speeds increased from 15.1 Mbps to 25.06 Mbps; while registered domain names grew from 75,096 to 93,130 (see Annex D).

Furthermore, the number of active monthly social media users rose from 7.7 million in January 2018 to 11 million in January 2021, increasing the social media penetration rate from 15% to 20.2% as shown in Table 2 in Annex D. Table 3 in Annex D shows that apart from WhatsApp, Facebook and YouTube command the highest overall usage with 9.5 million and 7.8 million active users in January 2021, respectively. Platforms such as LinkedIn, Instagram, Twitter, and SnapChat reported at 2.5 million, 2.3 million, 1.1 million and 1.3 million, respectively.

Social media in Kenya is mainly used for communication, expression, education, e-commerce, entertainment, and for social issues, e.g. news, politics, fashion, health, and gossip.[6] According to a US International University (USIU) study, the motivation for using social media includes acquiring information (31%), entertainment and pleasure (27%), social interaction (24%), personal identity (11%), and to escape some things (6%). Furthermore, the study found that YouTube was used for entertainment (74.4%), Facebook for entertainment (60%) and social issues (48%), WhatsApp for family (40%) and social issues (49.3%), and LinkedIn for job-related issues (61.9%) and education (42.1%). A 2021 study by ECPAT,[7] INTERPOL, and UNICEF found that for children aged 12–17 years, the main online activities included watching videos (57%), using social media (51%), using instant messaging (39%), playing online games (34%), watching live-stream (34%), school

work (25%), searching for new information (25%), and following celebrities and public figures on social media (20%).

Studies have found that social media platforms could be used in several ways to enhance and promote peacebuilding.[8] These include to mobilise for peace; influence peace negotiators; conduct peace campaigns and outreach; facilitate strategic communication and access information and updates including between mediators, groups in conflict, and wider public; facilitate dialogue, consultations, and building trust between conflicting parties; organise and mobilise social movements and political change; enable direct and unhindered reach to influence perspectives and engage the general public; enable different state and non-state actors to present their distinct narratives; facilitate data collection for conflict analysis and monitoring; counter misinformation and disinformation; bring in multiple voices into peace conversations; and complement ongoing face-to-face mediation and dialogue processes and initiatives. However, social media is being used to perpetuate biases, polarisation, violence, misinformation and disinformation, divide society, and exacerbate hate speech and related narratives.

Overview: Impact of content moderation on peace and stability

This section outlines the categories of potentially harmful content identified as prevalent on social media in Kenya such as hate speech, disinformation and misinformation, and online gender-based violence. It also discusses their impact on Kenyan society.

Problematic content in Kenya

According to Facebook, between April and June 2021, the single largest policy category that was actioned was fake accounts (1.7 billion), followed by spam (794 million). On Twitter, the major categories actioned between July and December 2020 included hateful conduct (31.9%), abuse/harassment (27.3%), and child sexual exploitation (13.3%), while 46.1% of the accounts were suspended for child sexual exploitation, 15.6% for hateful conduct, and 12.6% for impersonation. On YouTube, the leading reasons for the removal of videos included child safety (31.5%); violent and graphic content (19.9%); nudity and

sexual content (18.4%); spam, misleading and scams (9.1%%); harassment and cyber bullying (8.6%%); and harmful and dangerous content (8.1%).

Social media companies do not publish country disaggregated data on the prevalence of the categories of problematic content, volume of content removed, means of removal, types of flaggers, reasons for removal, number of appeals received or their outcomes, or how adverts are targeted. Consequently, it remains difficult to ascertain the prevalence of each type of problematic content and how community standards are applied in Kenya.

Hate speech: Hate speech[9] is linked to incitement to violence, propaganda for war, ethnic stereotyping, contempt and the use of abusive and derogatory language against marginalised communities, gender, sexual minorities, and persons with disability. Hate speech can also be rampant in political speech, especially when it seeks to polarise. Hate messages include those that spur ethnic hatred, discrimination, incitement to violence, attacks on gender identity, and use of stereotypes. During Kenya's 2007 election, the continued use of insults against opponents, threats of violence, incitement to violence, covert hate speech, defamatory and unsavoury language, propaganda, and political utterances were condoned, cheered on, and circulated by Kenyans through email, SMS messages, blogs, photos, publications, and media houses, especially vernacular language FM stations.[10]

During the 2013 and 2017 elections, hate speech and propaganda migrated from SMS and radio to email, online blogs, and social media platforms such as Facebook and Twitter. The Freedom House 2018 report on Internet freedom in Kenya notes that Cambridge Analytica was active during the August 2017 election through two websites that were used to spread hate speech and negative ads against the main opposition candidate, one of the sites spread positive narratives favouring the incumbent. Research has shown that social media platforms were used to propagate propaganda, tribal biases, ethnic mobilisation, incitement, and hate speech, which are feared to be a catalyst for predatory behaviour such as ethnic based mob-violence. A study by PeaceTech Lab on youth and radicalisation in Mombasa found a correlation between violent extremist activity and hate speech, and noted that social media alone did not advance violent extremisms, but was a component

in an inter-locking network that could drive individuals towards a path of radicalisation for violence.

According to a survey conducted in Nairobi in 2013, it seems that social media users are not sufficiently trained to report hate speech on social media. The survey found that while 54.4% of the respondents had taken part in peacebuilding initiatives on social media such as rallies and online events, only 17.8% were conversant with social media as a tool for reporting violence, although 45.6% had reported violence through police SMS hotlines or the Uwiano or Ushahidi Platforms. It also found that 63.3% of the respondents had received updates from groups spreading hate speech on social media, with 47.3% of them opting to comment and forward the messages to their friends online. In particular, the study found that the lack of ethical standards and professionalism of users contributed to the spread of unverified information, distorted facts, hate speech, and incitement that could be detrimental to peacebuilding and conflict prevention by igniting social tension.

Kenya's Constitution protects the right to freedom of expression in its Bill of Rights and stipulates that it does not extend to propaganda for war, incitement to violence, hate speech, or advocacy for hatred. As part of the measures to stem hate speech, the NCIC was established in 2008 to promote national cohesion, unity and reconciliation, and eliminate discrimination, including through receiving complaints and investigating complaints about potential cases of hate speech online, including on social media prohibited under section 62 of the National Cohesion and Integration Act. The NCIC has conducted peace campaigns and monitored hate speech since its establishment, especially during election periods. Other peace actors have also utilised social media platforms to call and campaign for peace during election periods, and for monitoring, documenting, and reporting on violence, tensions, and friction points.

Online gender-based violence content: A 2014 study revealed that while most Kenyans understood what gender-based violence was, the prevalence remained high with a lifetime prevalence[11] of 38% for women and 20.9% for men. Moreover, as of 2021, 40% of women in Kenya were likely to face physical and gender-based violence in their lifetime, including physical and sexual intimate partner violence. Violence offline is often replicated online. A

2016 study by the Africa Development Bank revealed that 33% of respondents, a majority of whom were women, had experienced online violence in forms that included personal hate speech, harassment, online intimidation, misogynist comments on social media profiles, cyberbullying and trolling, and sharing of sexual images without consent ('revenge porn'). Because of the attacks, the respondents indicated that they had either blocked the aggressor, exited the platform, or reported the abuse to the platform, with very few reporting to the police. A study by the National Democratic Institute found that insults constituted 71% of the online violence faced by women on Twitter in Kenya, which were often coupled with trolling to censor, embarrass, and inflict emotional harm on women.

In Kenya, women are not only attacked for their opinions, but also based on their gender, sexuality, and appearance. The attacks were also more pronounced during the Covid-19 pandemic, and where the specific women are prominent personalities and have presence in the public sphere, such as media personalities Yvonne Okwara and Janet Mbugua, and politicians such as Susan Kihika, Millicent Omanga, and Martha Karua who have all faced attacks online. According to a recent KICTANet study, online harassment of women has not gained attention from policymakers and the public, and outcries against harassment are viewed as rants from toxic feminists. Further, key populations including sex workers and LGBTQIA (lesbian, gay, bisexual, transgender, queer or questioning, intersex, and asexual) communities continue to face cyberbullying online.[12]

Misinformation, disinformation, and coordinated inauthentic behaviour: The prevalence of misinformation and disinformation has increased with respect to the governance and political speech (elections and political contests), public health (Covid-19 pandemic and vaccines), and climate change. A 2017 study by GeoPoll found that 90% of 2,000 people surveyed from 47 counties had seen or read false or inaccurate information during the 2017 electoral period, with social media being the dominant source. The same study found that 87% of the respondents were able to say that they saw false news, meaning they were able to identify false information. Notably, the study found that only a third of Kenyans felt that they were able to access accurate information about the election. The study also noted that while traditional channels such as radio, TV, and newspapers remained highly trusted sources of information, usage of social media was high but trust

in the accuracy of information found on social media remained lower compared to the traditional sources. This said, the competition for 'breaking news' led to a situation where the media shared stories on social media without necessarily first verifying the information.

Research also shows that the reasons why people share disinformation are complex: a 2021 comparative study notes that, while further research on disinformation in the specific contexts of sub-Saharan African countries would be needed:

> *"though the boundaries between satire used for political ends and malicious or misleading information may be nebulous, the long social history of such practices in Africa makes this an* important factor to consider. Given the entrenched role of satirical and humorous content in informal networks of media use in Africa, and the progressive uses to which these types of intentionally false – *albeit not misleading – content have been put, media users on the continent might be less resistant to sharing information that they know is untrue."*[13]

Social media has exacerbated disinformation. In 2020, a study on online political trolls in Kenya found the practice of circulating manipulated front pages of popular newspapers such as the *Daily Nation* and the *Star* newspapers in schemes orchestrated to deceive the public. The study also found a network of 484 accounts participating in coordinated conversations on Twitter and Facebook to spread disinformation themed against Kenya's Deputy President William Ruto. The campaign in May 2020 amplified trending Twitter hashtags such as #RutoGhostNumbers, #RutoWantedToKillUhuru, #RutoWantedToBetrayUhuru, #RutoTheWife-Beater, and #RutoMustGo by tweeting, retweeting, liking, replying, and mentioning each other's posts; it garnered 23,670 mentions and the hashtags were posted by 10,923 unique accounts in five days.

Another study by Code for Africa revealed a similar pattern of information manipulation using coordinated conversations through Twitter hashtags to spark conversations and spread misinformation in ways that trended about various political issues in the country in January 2021. The hashtags included #UhuruPettyLie, #ClassWarLoading, #ICCWatchList, #UhuruNMS27B, and #RutoReturnOurMoney, all of which received a total of 12,525

interactions on Twitter, comprising 1,643 tweets and 10,882 retweets from 4,616 unique accounts.

Likewise, a 2021 study by Mozilla Foundation on disinformation campaigns in Kenya found that malicious, coordinated, and inauthentic attacks on Twitter were discrediting and undermining the work of journalists, judges, and civil society in Kenya by muddying their reputations and stifling the reach of their messaging with a view to silence them. The study found that Twitter had no incentive to act and did little to curb the exploitation of its platform. It also exposed the lucrative and targeted disinformation campaigns that trended on Twitter, with 3,700 participating accounts spreading more than 23,000 tweets, and disinformation influencers reported being paid between USD 10–15 for participation in the shadowy campaigns. Moreover, the study found an industry of Twitter influencers with clear targets to exploit Twitter's features to sway public opinion by amplifying smear campaigns. After this study was released, Twitter acted on over 100 accounts which it found had violated its rules on manipulation and spam. One of the net effects of these disinformation attacks are the self-censorship, reputational damage, and enforced silence of some Kenyan activists on Twitter.

During the interviews conducted for this report, respondents observed that users had learnt to instrumentalise and weaponise social media to propagate problematic content and censor speech in several ways without detection or action by the platforms.[14] These include tactics such as setting up groups on social media specifically for sharing problematic content;[15] creating and coordinating troll armies or 'keyboard warriors' to run smear campaigns, threaten, intimidate, or coerce individuals;[16] taking down content by manipulating content reporting tools and alleging copyright infringement;[17] amplifying problematic content and making it go viral or trending to ensure maximum reach; and circumventing detection measures including by using multiple accounts, social media bots, and re-contextualising content to avoid detection by automated systems, e.g. posting in comment sections and across platforms such as WhatsApp.[18]

In September 2019, Edwin Mutemi wa Kiama's account on Twitter (@WanjikuRevolution) was suspended following a previous suspension in June the same year. The online activist's account was reinstated in February 2020. The suspension and take down resulted from a Digital Millennium Copyright Act complaint, which fellow bloggers also termed as malicious manipulation of the reporting system to promote censorship, an emerging practice in the country. According to media reports, the existence of the individuals who reported his posts on Twitter could not be independently verified, other than on their Twitter social media handles.

During the interviews conducted for this report, respondents noted that social media users often shared problematic content without regard for the consequences, accuracy, interpretation, or effect.[19] Even when disinformation was fact-checked and corrected, not many saw the corrected information, which engendered the idea that the initial set of misleading information was indeed true and fed into conspiracy theories about censorship.[20] Many times, the problematic content remained online.[21] According to the interviews conducted as part of this study, public awareness on fact-checking and identifying problematic content is still low, and many individual users have yet to build a culture and practice of fact-checking. Also, fact-checking organisations are few in Kenya (PesaCheck, Africa Check, and AFP Fact Check have been contracted by Meta) and may do not have the capacity to monitor the entire social media ecosystem and respond with fact-checked content in good time.

Escalating tension in social media spaces

According to respondents, online content is a by-product or a reflection of the society,[22] meaning individual biases, prejudices, and beliefs based on generations, sex, morality, culture, tradition, ethnicity, religion, origin, social background, status, or political affiliations can influence and shape the reasoning, arguments, and biases on online discussions and debates. According to a respondent, there is no such thing as a single standard of morality or common religion in Kenya. In addition, politics is a highly divisive subject which fuels tension between groups on opposite sides of the political divide.[23] In an increasingly

divided and polarised society, differences in perspectives can create and fuel tension over what content is deemed acceptable to stay up or problematic enough to be taken down. While social media by its very nature creates a connected global community, there is also a concern that it creates silos where users can interact with predominantly like-minded people, which could exacerbate polarisation and divisions and be a breeding ground for extremist views and hate speech. It is, however, not clear to what degree social media actually creates and reinforces so-called 'filter bubbles': research is ongoing, but there are signs that users, collectively, are exposed to a higher diversity of news sources on social media through 'incidental exposure' than in the offline context.

Another respondent stated that there was regulatory tension arising from the contest of who between platforms, governments, and users should determine and be the custodian of the community standards for regulation.[24] While some respondents observed that community standards could not fit all the needs and standards of culture, ethics, or morality across the country and the globe,[25] they emphasised the need for them to conform to universal human rights standards.

Widening discrimination of minorities and marginalised communities

Racial, ethnic, and religious minorities, LGBTQIA, persons with disabilities, albinos, and women are groups already affected by discrimination offline and are at risk of further discrimination online. They are disproportionately affected by entrenched stereotypes, prejudices, and discrimination in online spaces, thus reducing the diversity of voices online.[26] Respondents of this study observed that the platforms were not as sensitive to their needs and, as such, did not offer them an adequate level of protection against problematic content.[27] In addition, the respondents highlighted that ethnic minorities were often not heard, not present, or were silent in political debates dominated by larger communities, thus pointing to a bigger marginalisation problem.[28] However, some of the minority communities with access to social media also operated in silos in WhatsApp and Facebook groups, and, in some cases, also spread problematic content.[29]

Moreover, respondents noted that the growing digital divide had magnified the impact of the pre-existing marginalisation of communities that had been historically marginalised

(e.g. Turkana) and of digitally-excluded populations such as those living in informal settlements, rural areas or marginalised counties.[30] Respondents observed that these communities had limited access to digital devices, Internet, and social media platforms due to the high cost of devices and Internet access.[31] Some of these communities are not digitally literate, with most having feature phones, and are thus not present on social media platforms to articulate their opinions, narratives, engage in fact-checking, or report problematic content,[32] including content which often misrepresented the situation in their areas. Problematic content affecting them often remained online. In addition, the low literacy levels impeded their ability to read and understand the social media community guidelines and their ability to access the redress mechanisms[33] and the mechanisms for seeking redress from regulatory bodies, compared to those living in urban areas.

Moreover, exclusion was reported by respondents to be more pronounced among the elderly and persons with disabilities, a majority of whom are not on social media platforms.[34]

During the interviews conducted for this study, respondents pointed out that gender-based violence was often replicated on social media.[35] Furthermore, where women were seen as not conforming to the perceived or expected patriarchal roles of women, they were targeted, especially if they were public figures.[36] As discussed earlier about online gender violence, prominent women including journalists, TV presenters, and Members of Parliament have been attacked, trolled, and harassed, and are victims of toxic behaviour on Twitter, but little or no action is taken. Also, some women opt to stay away from social media after being attacked online.[37]

Similarly, the LGBTQIA community face online harassment, abuse, and trolling, making it difficult for them to speak, express themselves, and associate freely on social media.[38] A majority of them have practiced self-censorship, leaving only a few brave ones out in the open. Furthermore, respondents noted that LGBTQIA content was generally perceived as immoral and unacceptable to the Kenyan society, as can be illustrated by the decisions of national regulators such as the Kenya Film Classification Board, which has banned two

films with LGBTQIA content, namely 'Rafiki' and 'I am Samuel' for promoting sex, obscenity, nudity, LGBTQIA, and blasphemy contrary to Kenyan morality standards.

Issues with content moderation in Kenya

Removal of legitimate content

It has been observed that social media platforms enforce their policies in ways that seem too often inconsistent and opaque. Debates over content moderation decisions of social media platforms have been discouraged by a severe lack of publicly available data about the numbers or the reasons of those decisions taken by social media platforms. There is to date very limited independent oversight in place to check how they apply the rules, and their content moderation decisions. Furthermore, the platforms take down vast amounts of content and disable or suspend social media users' accounts based on rules, which they apply unevenly. The case study below highlights YouTube's contradictory content policies and practices.

This case study documents the experience of Sigi Mwanzia[39] in June 2021. She tried to share an advert on YouTube for two documentaries on Kenya[40] and Uganda[41] which were published in April 2021 on ARTICLE 19's YouTube Channel to highlight the report *Unseen Eyes, Unheard Stories: Surveillance, data protection, and freedom of expression in Kenya and Uganda during Covid-19*. The paid advert was rejected for failing to meet YouTube's Advertising Policies on 'sensitive events'.[42] Accordingly, YouTube suggested that Sigi either edits and resubmits the advert or appeals the decision. ARTICLE 19 Eastern Africa opted out of pursuing either option due to time limitations under the specific project. In her view, the disapproval of the advert was contradictory because the documentaries had already been uploaded and circulated on YouTube months before. This showed a discrepancy in YouTube's policies. The rejection of the adverts affected the work of the organisation and its objectives under the project by curtailing its ability to showcase and widely disseminate the informative documentaries to the targeted audiences within the region through YouTube.

One respondent observed that platforms would need to be careful about their approaches to ensure that their content moderation practices did not encourage users to a form of soft-censorship.[43] The same respondent suggested that platforms would seek to promote the diversity of opinion on the platforms and allow opportunities for a right of reply with knowledge and facts;[44] expand the space for discussion and not constrict it.[45]

In 2019, Twitter was criticised for its inaction on tweets from former US President Donald Trump and other leaders, which appeared to violate its content policies. The company defended its handling of posts from world leaders stating that blocking or removing their controversial tweets would 'hide important information people should be able to see and debate' and that 'It would also not silence that leader, but it would certainly hamper necessary discussion around their words and actions.' In a sudden change of policy in January 2021, Twitter 'permanently suspended' Donald Trump's account for violating its Glorification of Violence Policy. In June 2021, the company deleted a tweet by Nigerian President Buhari, to which the Nigerian Government retaliated by suspending the platform from the country. While sudden changes of content policies may be problematic, this study notes that the application of international human rights standards could provide a solid ground for changes and evolution in content rules.

Ineffectiveness of remedies

Ideally, all social media users should have access to internal mechanisms to address situations where their content is moderated. However, users face various challenges when they try and make use of such processes. The key challenges include the low levels of awareness of how to report problematic content or accounts, or access the appeal processes to restore content or accounts; basis and reasons for takedowns; a lack of technical know-how, knowledge, and capacity to use the internal complaint mechanisms, e.g. due to language barriers or digital inequalities; the lengthy response times for action on reported content or accounts; and the lack of trust of systems.[46] Thus, only those who are aware or familiar with the processes are able to complain. In addition, a 2016 study revealed that most users did not read the terms of networking websites before joining the platforms.

It should be noted that the content rules are not currently available in all local languages: at the moment, only the Facebook Community Guidelines are available in Kiswahili, while YouTube and Twitter rules are not.

Respondents noted that while platforms publicised the use of their services for entertainment, they do not place similar emphasis on reporting mechanisms.[47] Moreover, the community standards were neither prominently visible nor easily accessible within the main pages of the social media applications: on the contrary, information is tucked deep inside the apps. For example, on the YouTube Android mobile app, the Terms of Service can be accessed under the Settings menu, where they are displayed in fine print in the footer, from which a link to the community guidelines can be found at the bottom of the page. On the Twitter Android mobile app, they are found under the Help Centre menu under Rules and Policies. On the Facebook Android app they are under the Main menu > Help and Support > Terms and Policies > Community Standards, or under the Main menu > Settings and Privacy > Community Standards and Legal Policies > Community Standards. It is noteworthy that while the platforms are keen to get more people to comply with the community standards, keeping the standards almost hidden within the apps appears to be counter-productive.

Madina Chege, a Clinical Officer and social activist, uses Facebook (Madina wa Chege) to engage with her community of 7,737 followers. She helps people with advice on medical aspects, but also on governance and political issues. Her account was suspended for 30 days for sharing a video from the BBC YouTube page titled, 'China's Left-Behind 'Galamsey' Pikins'. In her view, the action was unfair and a move to target and silence her.[48] As a result, she became paranoid of posting material on the platform, and feared arrest. To restore her account and content, she had to reach out to platform representatives because the internal mechanisms failed her.[49] Lastly, some of her followers stopped following her page,and questioned her character, adding to the challenge she had of explaining the 'unfairness' of the suspensions.

Generally, the internal mechanisms were reported to be faceless, alien, unresponsive, not user-friendly, and consequently misunderstood by users.[50] Other respondents observed that platforms applied different rules with respect to notices, warnings, and appeals across the different applications.[51] According to one respondent, the escalation mechanism on Twitter was problematic because once the system recorded that an appeal had been made, no further intervention could be made by the user, which limited the possibility for an aggrieved user to get feedback, did not provide any alternatives such as contacts with human moderators, and did not even mention a timeframe for the platform to respond to the complaint.[52]

Moreover, a few respondents criticised the platforms as only doing the bare minimum in Africa, and were not investing adequate resources to guarantee the effectiveness of the complaint mechanisms for users in the region.[53] Indeed, the information revealed by whistleblower Frances Haugen has shown that the efforts of tech companies are far from being evenly distributed geographically. Most content moderation resources are being allocated to a limited number of countries.

Edwin Mutemi wa Kiama is a blogger and an online activist[54] on Facebook (Mwalimu Mutemi Wa Kiama/Wanjiku Revolution Kenya). His account was suspended and some posts taken down after sharing messages critical of government. One time, six of his posts were taken down in one day after they had been red-flagged. He managed to escalate the issue to Meta's Africa Policy team who acknowledged that the take down was erroneous and reinstated his posts.

According to respondents, being able to contact representatives of social media platforms directly seems the most effective way for users to report on errors in content moderation decisions and get a response from the platform. Yet, there is no transparency about who the representatives of social media platforms are in Kenya. It is also worth noting that it can take a lot of work and networking to find a personal connection to an individual who works for a social media platform in order to escalate a complaint and possibly get a

response.[55] Consequently, users who do not know how to reach platforms' staff through personal or professional connections find themselves in a worse situation.[56]

Group Kenya, which was one of the oldest and largest Facebook groups in Kenya with at least 2.1 million members, was deactivated in June 2020. The group was a platform for general discussions, news, opinion, politics, humour ,and entertainment. The 7-year-old group was suspended in June 2017 but was restored.

According to the group's administrators, the page was disabled for violating community standards on nudity. The administrators stated they disagreed with the assessment and that they would appeal the decision. Despite their appeal, the account has not been restored to date, and it is not clear whether restoration will be possible. Subsequently, the administrators created a new Group Kenya page which currently has 161,617 likes and 162,979 followers. A similar Facebook group under the same name exists with 242,000 members.

Individuals who have escalated to Meta or Twitter staff, indicated that without the intervention from the platform's representatives, they would not have succeeded in getting redress. However, even if such representatives are accessed, the interventions have not always been successful.

The case of Group Kenya shows that the decisions of social media platforms are not always well understood by administrators of such groups. Deactivating pages with large membership can be perceived as excessive, and consequently limiting the right and ability of the public to freely associate online. While problematic content can be spread in such groups or pages with a large following, actions taken by platforms need to be balanced to not only ensure due process, but also ensure that rights to expression and association are respected.

The study notes that platforms use terminology such as 'suspension' which ordinarily alludes to a temporary withholding of access to accounts, yet, in practice, some 'suspend' user accounts indefinitely without communicating the permanent nature of the said

suspension or the period, if temporary. For example, Twitter indicates that it may suspend a user account 'temporarily or, in some cases, permanently'. This can be confusing to users, who might believe that their suspension is indeed temporary, yet it could also mean that the said 'suspension' is indeed permanent and unappealable.

In May 2016, YouTube terminated NTV Kenya's YouTube Channel[57] following 'multiple third-party claims of copyright infringement regarding material the user posted'.[58] The company had received several warnings or 'strikes' as part of the US Digital Millennium Copyright Act. At the time, the channel had 50,775 videos, 200 million views and had 280,894 subscribers. Similarly, in October 2018, KTN News[59] Twitter account (@KTNNews) was suspended for violating the content rules, specifically for airing copyrighted content from the English Premier League.[60] The account had an estimated 470,000 followers and to date it has not been reactivated. The station now operates a verified Twitter account (@KTNNewsKE). While the two media enterprises may have violated copyright, the suspensions of the two accounts had implications on the freedom of the media. The actions of the platforms, while justifiable, ought to have been assessed and balanced against the importance of the accounts to Kenyans as critical news sources and freedom of the media; alternatively, a better option would have been the establishment or agreeing upon of a defined procedural mechanism/process for deleting the problematic posts.

Impact of automated mechanisms of content moderation

Automated systems are used by social media platforms to sort, index, curate, prioritise, or promote user content, as well as to detect and remove problematic content.[61] For example, Meta uses such systems to identify and remove content from Facebook. According to information shared by Meta, more than 90% of the problematic content are taken down through automated systems. During the Covid-19 pandemic, platforms increasingly relied on automated content moderation to tackle Covid-19 misinformation and disinformation.[62]

However, these systems have shortcomings such as being prone to bias and mistakes, such as removal of legitimate content.[63] These automated systems can perpetuate human biases, and a few human reviewers with insufficient contextual background and local knowledge may not be sufficient to rectify the errors. In addition, since the platforms' advert-driven business models are designed to maximise user engagement, the automated systems can amplify and spread problematic content because these types of content appear to trigger more engagement.

Additionally, it may simply be asked if platforms have enough human moderators to oversee automated decisions: generally, little information is available on how human content moderation is organised globally or per country or region, although recent reports have shed light on very harsh working conditions (see Presence of platforms in the country). Lastly, there is limited transparency and accountability on automated content moderation decisions, which undermines protections for freedom of expression and other human rights. While the level of transparency reporting has increased,[64] the companies still do not provide complete information disaggregated by country on the prevalence of harmful or problematic content, their enforcement decisions, or the role of automated systems in moderation.

Impact of global content moderation practices at the local level

Global rules and local context

The main basis for the moderation of content on social media are community policies, rules, guidelines, or standards. There is no uniform standard for content moderation, resulting in content moderation practices varying across social media platforms. Social media platforms also define, classify, and categorise the problematic content differently, and research has observed a lack of definitional clarity of problematic content such as hate speech. The Facebook Community Standards outlines categories such as violence and criminal behaviour, safety, objectionable content, integrity and authenticity, and respecting intellectual property. Twitter Rules has three broad categories, namely safety, privacy, and authenticity. The YouTube Community Guidelines provides six broad categories including spam and deceptive practices, sensitive content, violent or dangerous

content, regulated goods, misinformation, and monetisation (advertiser-friendly content guidelines).

While social media companies indicate that they comply with the laws of countries they operate in, the extent to which national laws such as the National Cohesion and Integration Act[65] and the Computer Misuse and Cybercrimes Act,[66] and directives from government officials such as the NCIC, Ministry of Interior, and Ministry of ICT, impact their decisions is not clear. Nonetheless, it has been reported that the government has 'increasingly sought to remove online content', including through regulation of content that it deems 'immoral' or 'defamatory'. Kenya has also considered a specific social media law, though it is yet to be enacted.[67]

Prior to the 2017 elections, the Communications Authority[68] and the NCIC issued guidelines to regulate political messaging on social media, including requiring social media service providers to pull down accounts used in disseminating undesirable political content within 24 hours. Also, section 62(1) of the National Cohesion and Integration Act[69] penalises media enterprises for the publication of speech that constitutes 'ethnic or racial contempt' while the Computer Misuse and Cybercrimes Act, 2018[70] in sections 22 and 23 penalises the dissemination and publication of false information.

How platforms such as Facebook enforce 'local laws' has been described as opaque. Meta indicates that between July and December 2017, it restricted access to 13 items alleged to have violated hate speech and election laws during the 2017 elections; in contrast to most of Meta's transparency reports, the information on content restrictions based on local laws is disaggregated by country. It is unclear the extent to which YouTube and Twitter enforce local laws as they have not reported on them. According to respondents from the NCIC and National Gender and Equality Commission,[71] the commissions have not engaged effectively with the platforms on content reporting and take downs. However, the lack of country-specific data on problematic content makes it difficult to assess compliance of the platforms.

Compliance of content rules with human rights standards

With respect to enforcement, some respondents reported that the platforms had challenges in striking a balance between promoting freedom of expression and preventing the dissemination of problematic content.[72] Notably, community standards have not always been defined or implemented based on international human rights standards.

Analysis by ARTICLE 19 in 2018 found that Facebook Community Standards, Twitter Rules and related policies and guidelines, and the YouTube Community Guidelines fell below international standards of freedom of expression. For instance, the global content rules on Facebook were found to have imposed overly broad restrictions on 'hate speech' than would be found in legislation or international standards. Twitter Rules were found to be 'difficult to understand' and imposed restrictions on 'violent extremism' that were inconsistent with applicable international standards. The YouTube Community Guidelines were found to fall short of international standards on freedom of expression in areas such as the restrictions on 'violent extremism' and 'terrorist content'. Overall, the studies found most of the rules imposed by the platforms to be 'broad in scope', leaving significant discretion to the companies in their implementation, and thus highly likely to lead to inconsistent application.

Some of these platforms have since revised their content rules to some degree, and they have also committed to respect and promote human rights. Notably, in March 2021, Meta launched its Corporate Human Rights Policy, which sets out the standards the company will strive to respect and apply across its apps, products, policies, programming, and approach to its business. Similarly, Google has a Human Rights Policy, which articulates its commitment to respecting and upholding internationally recognised human rights standards in its operations, including responsible decision-making around emerging technologies. Moreover, Twitter is committed to freedom of expression and privacy.

Presence of platforms in the country

Currently, Google, Meta, and TikTok have local presence in the country. Google opened its office in Kenya in 2007 and has an estimated 58 employees. The Kenya office is organised

as a separate entity from Google LLC and is run and operated by its commercial agent Google Kenya Limited, which handles marketing and sales of Google products and services in Kenya; however, all contracts with customers are made with Google LLC and Google Ireland.[73] Twitter has no local office or officers based in Kenya, apart from a single public policy official serving sub-Saharan Africa who is based in Ireland.[74] Meta has a public policy team of a few staff based in Nairobi who have been covering public policy issues within East Africa since 2018. The team is responsible for monitoring policy issues at the intersection of technology, social media, and market entry in East Africa; informing its agenda in the region; promoting the use of Facebook as a platform for citizen and voter engagement to policymakers and non-governmental organisations and political influencers; creating and implementing country and regional policy programmes; building coalitions to advance and support key policy issues; and communicating Meta's positions on public policy issues.

In 2020, TikTok set up its office at the Nairobi Garage. The move to open the office was part of an investment strategy and campaign to acquire talent and promote content creation on its platform, including by holding creator sessions and meetings to urge more creators to join the 'fun, cool short video platform'. The platform hopes to gain a competitive advantage by targeting popular creators to develop content for its largely Generation Z and millennial users.

With respect to human content moderators, Meta has a global team of 15,000, while YouTube has at least 10,000, and Twitter has 1,500. A respondent noted that the companies generally do not make public any information relating to their human moderators.[75] Hence, it is not clear how the companies allocate the roles or moderation tasks per country, the number of languages that moderators are conversant with, the specific issues they respond to, or where they are located. In 2019, it was reported that Meta planned to set up a content review centre in Nairobi in collaboration with Sama, with an initial staff of 100 people.[76] While little information is available on Meta's relationship with Sama, a recent report exposed problematic practices at Sama in Kenya, which revealed that content moderators were resigning due to poor pay and mental illnesses, such as post-traumatic stress disorder, anxiety, and depression.

Relationships between stakeholders and platforms

Social media platforms rely on a combination of user reports, Trusted Flaggers, human reviewers, and automated tools to identify content that violates their policies and to moderate content. It is worth noting that as the volume of content grows, the platforms rely more on automated systems for moderation. Individual users can also report content that they deem to be harmful, inappropriate, or in violation of a platform's community standards. Indeed, all the major social media platforms, including Facebook, Instagram, YouTube, Twitter, and SnapChat have put in place such reporting mechanisms.

Trusted Flaggers may also initiate reports on problematic content. The 'Trusted Flagger status' is granted by social media companies to organisations that are deemed to possess proven expertise in evaluating violations of content policies. The status provides a direct communication channel to a social media platform's review team. For instance, in the case of YouTube, individual users, non-governmental organisations, and even government agencies can be granted such status, and content flagged by them are prioritised for review.

In Kenya, Watoto Watch Network is a Trusted Flagger for child sexual abuse content and collaborates with Google and Meta to promote child protection.[77] Likewise, the Bloggers Association of Kenya (BAKE) is a Trusted Flagger on Facebook.[78] These organisations were approached by the platforms to join the Trusted Flaggers programmes. Another form of collaboration between platforms and stakeholders is illustrated by the fact that PesaCheck, Africa Check, and AFP Fact Check are organisations contracted by Meta to fact check information: their contribution is used by Meta to flag content as unreliable.

There is, however, a lack of transparency around the Trusted Flagger programmes, including on how to join or enlist, the list of the Trusted Flaggers per country, and the categories of content they respond to, the actions taken by the platforms as a result of reports, and the content removed based on reports.

Meta has collaborated with different organisations on areas such as child online protection, digital safety, Internet governance, hate speech, disinformation, and so on.

Some organisations reported long-standing and valuable collaboration with Meta, such as Watoto Watch Network for their annual Safer Internet Day,[79] KICTANet for the Kenya Internet Governance Forum, Kenya School of Internet Governance,[80] Kenya Editors Guild for their convenings,[81] and PesaCheck who are contracted to fact check content on its platforms.[82] Some of the social media platforms were involved in the development of guidelines on bulk messaging by the NCIC and the Communications Authority.[83]

Further, according to respondents, Meta and Google have had ad hoc campaigns prior to and during election periods to tackle hate speech on their platforms.[84] The Facebook Journalism Project and Reuters launched a free e-learning programme to train journalists on digital news gathering, news verification and reporting, publishing on social media, wellness, and resilience training while reporting. Twitter has no reported engagements with stakeholders in the country and it is not clear why. Overall, the platforms are not transparent about their engagements with the governments in the country[85] or the organisations they provide financial support to.

Facebook's Oversight Board

To date, only Meta has constituted an external oversight complaint mechanism known as the Oversight Board. As of October 2021, the Board has adopted 18 decisions, of which none relates to a case from Kenya. Out of the 524,000 cases submitted between October 2020 and June 2021, only 2% were from sub-Saharan Africa. The Board admits that it does not believe that the outcome 'represents the actual distribution of Facebook content issues around the globe'.

Maina Kiai, a Kenyan, is a member of the Board and could provide an in-depth insight of both the local and regional context to the Board. On his appointment, he welcomed cooperation with local stakeholders and encouraged the public to raise issues for investigation by the Board.

Challenges of local context

While social media platforms use the word 'community' to refer to their users globally, one respondent suggested that the 'community standards' should be negotiated with real-

world communities instead of being a 'one size fits all' document.[86] According to another respondent, elaboration and enforcement of global rules appears to be implemented without due regard for the needs of 'communities'.[87]

The interpretation of policies does not seem to take the cultural context into account. On Instagram Community Guidelines for instance, nipples of men and women are treated differently, as the platform prohibits only photos of female nipples,[88] and permits those of males. However, among the Turkana, since women and girls culturally wore no tops and, when photographed, their images are taken down on Instagram for violation of community guidelines.[89] In addition, some cultural practices such as animal slaughter were often taken down for violating community standards for not being animal friendly.[90]

A respondent pointed out that decisions on content moderation were made by social media platforms officials based in the US who do not always have an understanding of the local knowledge, language, and context.[91] Another respondent observed that while companies such as Meta had deployed a few public policy representatives within East Africa, some like Twitter only has single public policy official serving sub-Saharan Africa – and that person is based in Ireland.[92] This could contribute to the lack of understanding of the local context and inaction on key issues in the country.[93] Moreover, there is limited research on content moderation in Kenya and support for such research and work is equally limited.

According to respondents, the automated content moderation systems were unable to detect problematic content in local languages or detect the nuances or lingo within these local languages, which changed depending on the circumstances and/or region.[94] An abusive word in one language might not be abusive in another as in some cases the same words have different meanings in different languages,[95] and it can also happen that the meaning of certain words changes depending on the issues. For example, the phrase 'nitakufinyaa' (I will squeeze you) had acquired a new meaning and could be interpreted in many different ways in 2021 compared to how it was interpreted in 2020.[96] The effect of all this, according to a respondent, was that problematic content posted in local languages remained undetected by the automated content moderation systems, unless flagged or

reported by users.[97] Moreover, a respondent observed that little investment of platforms in moderating content in local languages as in other regions could be interpreted to mean that the needs of non-English speakers are not a key priority for them.[98] In the view of some respondents, the consequences of not taking the local context into consideration were that inaccurate information was not always removed from the platforms and that the marginalisation of certain groups was further exacerbated,[99] including by instrumentalising prejudices against some ethnic groups.[100]

Discrimination in content moderation practices

Interviewees have mentioned that there have been reports of discriminatory approaches by platforms in the moderation of content in countries in the Global South and the Global North[101] and that platforms are perceived as not doing enough in the Global South to tackle problematic content, including in situations where such content could undermine electoral processes.[102] Another challenge highlighted by respondents is that the platforms amend and enforce policies 'willy nilly', with limited engagement with stakeholders in the Global South.[103]

Leaked documents in October 2021 revealed that Facebook classified countries into tiers, which determined their approach to content moderation, including the tools, resources, and staff deployed. The leak showed an opaque system of evident inequality as the company lacked misinformation and hate speech classifiers[104] in some countries, and only placed language experts in select countries. Furthermore, the revelations by whistleblower Frances Haugen showed that Meta officials were aware of their sites' flaws and potential for harm but chose to put profits over the well-being and safety of users. The same revelations also show that the company was reluctant to apply a systematic approach to restrict features that disproportionately amplified incendiary and divisive posts. Moreover, it appears that the automated systems had minimal success in removing hate speech, violent images, and other problematic content. Indeed, Facebook's response to the Christchurch killings by taking down 1.5 million videos within 24 hours demonstrated its capacity to act fast and decisively: it means that they could apply the same tactics and do more to address problematic content in Kenya.

Interim conclusion

Content moderation has been described by a respondent as 'broken, ineffective, and slow' since 'where it was most needed, it was least used'.[105]

Respondents also insisted that it is critical for stakeholders to hold platforms accountable for their human rights impact, including on their transparency in complaints handling,[106] flagging of content from the region,[107] handling content affecting marginalised and vulnerable groups,[108] and the measures in place to promote awareness of the content policies and internal complaints and reporting mechanisms.

This study noted that a more sustainable and transparent engagement with local civil society organisations and other local stakeholders could be a way for platforms to reinforce their capacity to integrate a robust analysis of the local context (including the cultural, social, political, linguistic, and political dimensions) into their content moderation practices. This is particularly important in multilingual, culturally diverse, and ethnic-polarised countries like Kenya.

To address these issues, ARTICLE 19 and other civil society organisations have developed recommendations based on international standards on human rights.[109] Of particular importance for this study, is the principle of Culture Competence set forth in the Santa Clara Principles, which:

> "*requires, among other things, that those making moderation and appeal decisions understand the language, culture, and political and social context of the posts they are moderating. Companies should ensure that their rules and policies, and their enforcement, take into consideration the diversity of cultures and contexts in which their platforms and services are available and used (…), and companies should ensure that reports, notices, and appeals processes are available in the language in which the user interacts with the service, and that users are not disadvantaged during content moderation processes on the basis of language, country, or region*."

In addition, flaws in content moderation identified in this study should be addressed through the following recommendations:

- Companies should ensure that their content rules are sufficiently clear, accessible, and in line with international standards on freedom of expression and privacy. It is of key importance that social media companies' content rules be made accessible and available in local languages.
- Companies should also provide more detailed examples or case studies of the way in which their community standards are applied in practice and conduct reviews of their standards to ensure human rights compliance.
- Companies should be more transparent about their decision-making processes, including the tools they use to moderate content, such as algorithms and Trusted Flagger schemes.
- Companies should ensure that sanctions for non-compliance with their Terms of Service are proportionate.
- Companies should put in place internal complaints mechanisms, including for the wrongful removal of content or other restrictions on their users' freedom of expression. In particular, individuals should be given detailed notice of a complaint and the opportunity to respond prior to content removal. Internal appeal mechanisms should be clear and easy to find on company websites.
- Companies should publish comprehensive transparency reports, including detailed information about content removal requests received and actioned on the basis of their Terms of Service, including on a per country basis. Additional information should also be provided in relation to appeals processes, including the number of appeals received and their outcome.
- Companies should collaborate with other stakeholders to develop new independent self-regulatory mechanisms, such as a Social Media Council, modelled after effective self-regulation archetypes in the journalism field.

Analysis of stakeholders

This chapter describes the various stakeholders that work on aspects of online expression and content moderation in Kenya and puts forward observations on the roles they could play in the constitution and operation of a multi-stakeholder Coalition on Freedom of Expression and Content Moderation in Kenya. The list of stakeholders has been established on the basis of input and close consultation with UNESCO Regional Office for Eastern Africa.

Public sector actors

Key actors

- Communications Authority (CA)
- Independent Electoral and Boundaries Commission
- Judiciary of Kenya
- Kenya Copyright Board
- Kenya Film Classification Board (KFCB)
- Kenya National Commission on Human Rights (KNCHR)
- Media Council of Kenya (MCK)
- Ministry of ICT, Innovation and Youth Affairs
- Ministry of Interior
- National Cohesion and Integration Commission (NCIC)
- National Communications Secretariat
- National Cybercrimes Coordination Committee (NC4)
- National Gender and Equality Commission (NGEC)
- National Police Service – Cybercrimes Unit
- National Steering Committee on Peace Building and Conflict Management
- Office of the Attorney General (OAG)
- Office of the Director of Public Prosecutions (ODPP)

Flagship projects and initiatives

The Uwiano Platform for Peace was formed in May 2010. It is coordinated by the National Steering Committee on Peace Building and Conflict Management and brings various state

and non-state actors to provide leadership around political and electoral processes and a forum for improving coordination and linkage for electoral violence reduction. In 2020, the NCIC unveiled a roadmap to peaceful elections with an objective to set the agenda and direction for all peace actors to make adequate preparations for a peaceful electoral process.[110] The roadmap includes monitoring of social media for hate-related messages and infusing positive messages on social media through positive peace campaigns.

Capacity and needs

Ministries, departments, and agencies could be partners at the appropriate stages. Notably, commissions such as the KNCHR, the NGEC, and the NCIC are critical and have national reach with a capacity to support such an initiative; conduct public awareness and engagement; regulate and oversee; support development of policies and laws; coordinate actors; secure budgets to support initiatives within their mandate; and monitor compliance with human rights standards. Their key needs include capacity-building and sensitisation on content moderation, including on new approaches to regulating problematic content on social media platforms, including through self and co-regulation.

Risks and opportunities

The approach by the executive to content regulation has been problematic, as it is perceived to take the form of censorship.[111] Examples include the Computer Misuse and Cybercrimes Act, 2018 that enables censorship in the guise of tackling problematic content.[112] In addition, there are multiple institutions overlapping mandates on digital content regulation[113] without a coordinated approach to regulation.[114] Some respondents feared that some government agencies within the executive tended to exclude non-government stakeholders in engagement. However, the Communications Authority, KNCHR, NCIC, and NGEC were noted to be inclusive in their approaches.

Civil society organisations

Key actors

- Access Now
- Act!

- Africa Check
- Amnesty International
- ARTICLE 19 Eastern Africa
- Africa Centre for People Institution and Society
- Bloggers Association of Kenya
- Co-Creation Hub/iHub
- CEMIRIDE
- CHRIPS
- Civil Society Reference Group
- Code for Africa (iLab)
- ELOG Kenya
- FIDA Kenya
- Gay and Lesbian Coalition of Kenya
- Haki Africa
- HIVOS
- ICJ Kenya
- ICNL
- Katiba Institute
- Kenya Human Rights Commission
- KICTANet
- Kituo cha Sheria
- Law Society of Kenya
- Lawyers Hub
- Mozilla Foundation
- Open Society Foundation (OSIEA)
- Ushahidi
- The Elephant
- Peace Tech Lab
- Peace and Development Networks Trust
- Kenya Partnership for Peace
- PesaCheck
- Power 254
- Search for Common Ground
- Sentinel Project
- Ushahidi

Flagship projects and initiatives

There are few initiatives such as trainings on digital rights, fact-checking, and misinformation. Amnesty International, ARTICLE 19 Eastern Africa, and KICTANet, are members of the Africa Internet Rights Alliance coalition, which promotes digital rights.

Likewise FIDA Kenya, Act!, and ARTICLE 19 Eastern Africa are members of the Civil Society Reference Group. The Elephant publishes relevant digital media content on key topics on culture, politics, and society. Other organisations work on elections, conflict, and peacebuilding initiatives.

Capacity and needs

Civil society organisations interviewed expressed an interest in participating in a local coalition as some already work on digital rights, Trusted Flaggers, fact-checking, public interest litigation, and policy and legislative advocacy. Many supported ARTICLE 19's leadership of such a coalition. Civil society organisations generally have medium influence, but their impact within the country is high. Existing coalitions, networks, or working groups work on diverse human rights issues such as peace (Peace Net), defending civic space (Civil Society Reference Group), digital identity (National Integrated Identity Management System Coalition), freedom of information (FOI Network), elections observations and monitoring (Election Observation Group), etc. While the knowledge and skill levels varied, most of the respondents from digital rights civil society organisations demonstrated a medium- to high-level of understanding of content moderation. Civil society organisations could benefit from greater coordination through a coalition of the relevant actors.

Risks and opportunities

Civil society organisations generally work well together, but their political power and influence varies depending on their size, focus area, national reach, and geographic location. The national-level civil society organisations are generally more dominant, skilled and well-resourced compared to grassroots civil society organisations, thus capacity-building and inclusivity will be key. Civil society organisations compete for funding from the same sources, which could impact their work. There are few civil society organisations working on content moderation, and the overall work in this area is fragmented[115] as there is little collaboration and information sharing. Some of the recent initiatives are ad hoc or temporary. In addition, there is limited technical capacity, tools, and advanced techniques

to monitor how content is moderated online, and, as a result, there is limited data of the trends in content in the country.[116]

Civil society organisations remain central to social media and peace given their initiatives on positive cyber-citizenship, digital rights, democratic consolidation, and peacebuilding. Respondents from civil society organisations indicated their willingness to support a coalition and promote public awareness, capacity-building, advocacy, engagement, and fundraising. The absence of a coalition presents an opportunity to harness the diverse expertise of civil society organisation for greater impact. Further, strong donor support could be instrumental to the success of the coalition.[117]

Social media companies

Key actors

- Google (YouTube, Blogger)
- Meta (Facebook, Instagram, WhatsApp)
- Microsoft (LinkedIn)
- TikTok
- Twitter

Flagship projects and initiatives

Some companies such as Meta have engaged their partners, e.g. PesaCheck, to conduct fact-checking and collaborated with Watoto Watch and KICTANet to host events on topical issues. The company has also convened specific events with stakeholders on different issues relevant to its work. The support from the company is not viewed by civil society organisations as undermining their independence, but rather as partnerships to address mutual concerns within society, drawing from the shared expertise.

Capacity and needs

Respondents from civil society noted the need for platforms to be involved and engaged in the work of the coalition, but not for them to lead the initiative, given their very direct interest in the issues. Notably, social media companies are large and complex entities with diverse internal staff who have various sets of skills and are driven to achieve different sets of objectives. Their actions have significant impact since they own the platforms and determine the community guidelines and standards that are applied on the platforms. Social media company staff are also influential as they articulate company positions in the country, but also give feedback from stakeholders to the company.

Respondents observed that platforms could not tackle problematic content by themselves and would need to be more entrenched in the countries and work closely with local stakeholders, including researchers and civil society.[118] In addition, the companies need to be transparent and share more disaggregated data on the enforcement of community standards; the prevalence of problematic content; user perspectives on content moderation practice; and technologies used to track abuse. It will also be useful for the social media companies to have a more robust engagement in the region and work collaboratively to reinforce their understanding of how the global content rules should be applied in the context of Kenya.[119]

Risks and opportunities

The companies have high power and influence in content moderation given the fact that they own the platforms. As businesses, they are primarily motivated by profits generated from their advertising sources arising from social media content. Their willingness to participate will likely depend on their local presence and interest in stakeholder engagement. However, the authority and influence of locally based representatives could be limited. Locally, the companies are perceived by respondents to place more emphasis on their business interests, which supersede human rights and concerns, including tackling problematic content on their platforms.[120] Others stated that platforms appeared to still have challenges in respecting freedom of expression standards.[121]

Moreover, respondents have blamed platforms for not engaging effectively with users. For example, some respondents observed that Google/YouTube did not engage much with creatives, especially regarding Digital Millennium Copyright Act complaints. Likewise, there was little engagement by Twitter with stakeholders to act on hate content and disinformation.[122] Furthermore, the platforms did not generally seem to have a clear position on the responsibility for the effects of what happens on social media, which gives room for governments to shut down the Internet and block social media.[123]

Academia

Key actors

- Catholic University of Eastern Africa
- Daystar University
- Hekima University
- Maseno University
- Moi University
- Rongo University (Center for Media, Democracy, Peace and Security)
- Strathmore University (Centre for Intellectual Property and Information Technology Law, CIPIT)
- University of Nairobi (School of Journalism/Law)
- USIU Social Media Lab (SIMElab)

Flagship projects and initiatives

Institutions such as Strathmore University's CIPIT and USIU SIMElab have published research on social media, including on hate speech and disinformation. Hekima University's Institute of Peace Studies and International Relations (HIPSIR) had convened the Dialogue Contact Group after the 2017 elections, but it is not clear how the process has unfolded.

Capacity and needs

While there is limited research on content moderation in Kenya, there is potential for research and engagement, as shown by research by Strathmore's CIPIT and USIU's SIMElab. In addition, HIPSIR has programmes and research on peace and conflict. While

academic institutions have low political power, the impact of their work is high, which could promote awareness and inform policy and legislation on content moderation. There is also a need for continued funding for research, coordination, and awareness on content moderation.

Risks and opportunities

There are few institutions working on content moderation and the links with peace and conflict. Academia can monitor trends, provide thought leadership, and identify new approaches to tackle content moderation challenges.[124]

Media

Key actors

- Africa Uncensored
- Association for Media Women in Kenya (AMWIK)
- Baraza Media Law
- Kenya Correspondents Association
- Kenya Editors Guild
- Kenya Union of Journalists
- Media Owners Association
- Media Sector Working Group
- Nation Media Group
- OdipoDev
- Parliamentary Journalists' Association
- Royal Media Group
- Standard Media Group

Flagship projects and initiatives

Some organisations such as the Kenya Editors Guild, the Kenya Correspondents Association, and AMWIK have conducted training for their membership on fact-checking, misinformation, and online violence.[125] Most of the media enterprises moderate content on their social media platforms based on their editorial policies and in compliance with existing laws and standards.

Capacity and needs

The media's role in contributing to cognitive, attitudinal, and behavioural change on a large scale is unique. Media enterprises and associations are influential and could be allies in shaping public opinion, outreach, and awareness both offline and online. The Royal Media Group, Nation Media Group, and the Standard Group have significant political power given their relationships with Kenya's ruling political class. The Kenya Editors Guild and the Media Owners Association are also influential in the sector. The journalists' unions and associations have a significant following among individual journalists.

Risks and opportunities

Media enterprises are likely to support initiatives that align to their business and political interests. Engaging the associations could bring more diversity of membership, irrespective of political or commercial interests. The key needs include capacity-building on content moderation, fact-checking, and coordination.

Private actors

Key actors

- Kenya Private Sector Alliance (KEPSA)
- Public Relations Society of Kenya
- Safaricom PLC
- Sama Source
- Technology Service Providers Kenya (TESPOK)

Flagship projects and initiatives

KEPSA sponsored the MKenya Daima (My Kenya Forever, in Swahili) campaign, which promoted peaceful elections in 2013 and ensuring smooth transition. KEPSA, Media Owners Association, and Kenya Association of Manufacturers also spearheaded the conduct of presidential and gubernatorial debates in 2017. In 2013, Safaricom donated 50

million text messages to the Sisi Ni Amani (We Are Peace) coalition, which were used to promote peace. Mobile network operators, together with regulators and other stakeholders, developed the *Guidelines on Prevention of Dissemination of Undesirable Bulk and Premium Rate Political Messages and Political Social Media Content via Electronic Communications Networks* which regulate the use of social media for political content.

Capacity and needs

These organisations have capacity to engage in the coalition despite their varying knowledge of content moderation. KEPSA has significant influence and has collaborated with several stakeholders on peace, conflict, transitional justice, and addressing hate speech especially during pre- and post-election periods. Sensitisation on content moderation on social media and its implication in peace and conflict work will be a key need.

Risks and opportunities

Safaricom, Technology Service Providers of Kenya (TESPOK), and KEPSA are influential and close to government with respect to policy and legislative advocacy, although they have not specifically focused on content moderation on social media. Sama conducts content moderation for Meta.

Religious actors

Key actors

- All Africa Conference of Churches (AACC–CETA)
- Catholic Justice and Peace Commission (CJPC)

- Council of Imams and Preachers of Kenya
- Evangelical Alliance of Kenya
- Hindu Council of Kenya
- Inter-Religious Council of Kenya (IRCK)
- Kenya Conference of Catholic Bishops
- Kenya Episcopal Conference
- National Council of Churches of Kenya (NCCK)
- Supreme Council of Kenya Muslims

Flagship projects and initiatives

There have been several initiatives on hate speech, peacebuilding, and conflict such as the Dialogue Reference Group. In addition, the NCCK and CJPC have conducted peacebuilding and conflict transformation work in collaboration with the National Steering Committee on Peacebuilding and Conflict Management.

Capacity and needs

Some of these organisations have capacity to engage in the coalition given their work on peacebuilding, conflict, transitional justice, and addressing hate speech especially during pre- and post-election periods. While their knowledge of content moderation varies, they have significant reach and influence and could be useful avenues for promoting awareness on content moderation. Sensitisation on content moderation on social media and its implication in peace and conflict work will be a key need.

Risks and opportunities

The influence of these institutions can be an asset where there is shared understanding, and a liability when not. Approaches to promoting peace and preventing hate speech have been reactive, inconsistent, short-term, and focused on election periods, which is unsustainable. Previous peacebuilding campaigns were perceived as promoting the status quo and government positions, rather than seeking holistic measures to promote human rights, justice, and peace collectively. Likewise, some religious organisations have shown bias based on ethnic and political lines, thus affecting engagement and their legitimacy.

Conclusions

As social media use increases in the country, tackling the spread of disinformation and misinformation, hate speech, online gender-based violence content, and malicious, coordinated, and inauthentic behaviour remain key challenges not just for the platforms but for everyone. This study has documented the current flaws that affect content moderation practices on main social media platforms, such as a lack of country-level data on content moderation; algorithms that prioritise and amplify extreme, divisive, and polarising content; low public awareness and limited access to content rules in local languages; ineffective complaint mechanisms and remedies; marginalisation and exclusion of communities; lack of consideration for the various dimensions of the local context in content moderation practices; and inconsistent application enforcement of content rules.

In Kenya, significant attention is yet to be paid to address problematic content and behaviour on social media platforms given their implications on peace and stability in Kenyan society. The state and non-state stakeholders working on conflict, peacebuilding, digital rights, elections, and technology need to collaborate widely to address the impact of social media online and offline. As the 2022 election approaches, the spread of problematic content on social media is likely to rise, which provides an opportune moment to reflect on how social media companies and all relevant stakeholders could address the challenges highlighted in this study.

This study has shown that a sustainable and open engagement with local stakeholders could help social media companies to integrate a stronger understanding of the various dimensions of the local context into their content moderation systems, which would, in turn, improve the moderation of problematic content that negatively impacts the Kenyan society. Such approach would contribute to social media being a public good and a platform for a peaceful and democratic society.

Most of the respondents interviewed welcomed the idea of a multi-stakeholder coalition to work on the issues outlined in this study. In their view, such a coalition could be a useful platform to organise, engage, and co-create local, strategic solutions and responses to

tackle the spread of problematic content on social media in Kenya. In addition, it could provide a useful avenue to start engagement with social media companies and to promote international human rights standards, and transparency and accountability in content moderation.

Recommendations

The following steps are presented for consideration as part of a strategy towards the creation of a **Coalition on Freedom of Expression and Content Moderation** with the organisations listed in Annex B.

1. **Identify or be the strategic coordinator or convenor.** This requires a committed vision-bearer to coordinate and oversee the coalition. ARTICLE 19 can carry out this role directly, or with UNESCO as the coordinator and secretariat. Alternatively, they could designate a specific individual or organisation to coordinate the coalition's secretariat, and to facilitate communication and branding, outreach and stakeholder engagement, logistics, documentation of meetings, and finance.
2. **Determine the coalition model.** The coalition should be based on an opt-in model rather than consensus to ensure the independence of the coordinating body in decision-making, while at the same time permitting consultation and collective action. This model prevents the dilution of goals and strategies and allows the coalition to be dynamic to engage with multiple actors and coordinate its affairs, including pursuing specific objectives.
3. **Develop constitutive documents.** A coalition charter should articulate the coalition's principles, purpose, goals (short, medium, and long-term), objectives, values, and vision and mission. It could also include the proposed scope of work, a clear theory of change, examples of activities, and criteria for membership, structure, leadership (steering committee), decision-making processes, working methods, reporting, communication, and codes of conduct.
4. **Secure funding.** The coalition could be funded initially through seed funds or in-kind support to facilitate the establishment of the coalition, and outreach work to potential coalition members. Once formed, the member organisations could support the coalition either directly by implementing its activities or indirectly by sponsoring activities. In addition, the secretariat could seek donors for long-term unrestricted funding for strategic organisational support.

5. **Conduct outreach and recruitment to the initial core members.** These organisations can be reached out to individually or recruited through a common meeting. A tentative list of organisations has been provided in Annex B.
6. **Get commitments/onboard core members.** Those organisations or individuals who express an interest to joining the coalition should sign a Memorandum of Understanding (MoU). A database of each organisation, the contact persons, areas of work and expertise, ongoing work, member needs, and priorities should be collected and stored.
7. **Hold inception meeting or conference.** This convening should provide an opportunity for the members to meet based on a clear agenda, including the review of issues and the problems; the structure of the coalition and decision-making processes and procedures; the mission, vision, goals, and objectives to ensure they are SMART (specific, measurable, achievable, realistic, and timely); the member rules and codes of conduct; the action/workplan, and a roadmap for the coalition. They could also appoint the steering committee; designate or be introduced to the secretariat; be appraised of the available financial, material, or other resources to the coalition; and consider and agree on working processes, including communications, documentation, frequency and location of meetings, and branding.
8. **Announce the coalition.** Once a critical number of members have joined, the coalition can be announced to the public.
9. **Convene follow-up meetings.** Follow-up meetings can be held to review commitments from the initial meeting, and to update and conclude pending matters from the inception meeting. Such meetings remain important for information sharing, continued consensus and building of trust, strengthening relationships, and entrenching the shared vision and goals.
10. **Undertake capacity-building.** The key areas include content moderation issues and challenges, leadership and teamwork, data analysis, advocacy, and coalition building. These could be facilitated through regular stakeholder meetings, feedback seminars, platforms for facilitated discussion, and collaborative learning processes.
11. **Identify entry points.** The scope of the coalition could initially be limited to a few key priority issues and be expanded progressively based on the changing context given

the upcoming election in 2022. Key priority issues could include promoting stakeholder awareness on content moderation; conducting stakeholder engagement with platform representatives; and strengthening the coalition. The coalition can leverage on the work on identified stakeholders already working on elections, hate speech, and disinformation online.

12. **Develop joint advocacy, communication, and engagement strategies.** The coalition could develop joint advocacy, communication, and engagement strategies for outreach to government, private sector, donors, civil society, etc. The approaches should be inclusive and multi-stakeholder to ensure diverse, representation, and participation by the stakeholder groups, while paying attention to the needs of women, marginalised groups, youth, vulnerable groups as rights defenders, political dissidents, minorities, and indigenous communities. These strategies could ensure proper prioritisation, build trust, resolve diverse perspectives especially on contentious topics, and ensure alignment. The coalition will need to determine how to position itself vis-à-vis the government and the platforms. While the private sector and religious groups work more closely with government, civil society are not as close. Thus, engagement with government and platforms will need to be constructive, but also with caution.
13. **Develop action plans.** Develop and implement an action plan with clear performance targets and success indicators within the specified timeframes.
14. **Monitor, evaluate, learn, and report.** Document relevant incidents; monitor social media platforms; conduct regular research, assessments, surveys, interviews; and hold discussions with key stakeholders to obtain critical feedback, identify new issues, and to monitor and evaluate progress, feelings of members, and capture lessons that continue to emerge based on the work of the coalition. Regular monitoring could help document progress, provide evidence of impact, distil insights and lessons that could benefit, and improve and refine the coalition's strategies and processes moving forward. To ensure transparency, the coalition could publish annual reports of its work and the progress made towards the achievements of its objectives.

Annex A: Risk analysis

The **Coalition on Freedom of Expression Online and Content Moderation** emerges as a unique opportunity for participation and contribution by all the actors and as a mechanism for meaningful change. The coalition offers a path to consensus on key content moderation issues – and opportunities to address them. The following table provides an overview of the potential risks related to the formation and functionality of the coalition, identified by the respondents, including potential ways to overcome and mitigate them.

Risk type*	Description of risk	Likelihood**	Impact***	Monitoring and mitigation
Finance	The source of funds could affect the independence of the coalition	Possible	**Severe**	• Determine the sources from which funds can/not be taken; reinforce independence, and develop arm's length mechanisms to address funding and independence issues
Finance	Competition among members for funding or in managing funds	Possible	**Major**	• Have a policy on fundraising and conflict of interest; coordinate donors interested in supporting the areas to ensure better support for initiatives
Finance	Payment or non-payment of honoraria	Possible	**Moderate**	• Provide a basis for compensation and allow members to have an option to be paid or volunteer
Finance	Lack of long-term, sufficient and sustainable funding for the coalition's activities	Possible	**Severe**	• Have a fundraising strategy and a sustainable funding structure for the secretariat that is diverse, e.g. project or long-term unrestricted core funding, and member contributions to a pool or to activities in the workplan
Finance	Poor financial structures leading to disagreements, embezzlement, and mismanagement of the coalition's funds	Possible	**Major**	• Have financial policies and procedural policies in place; ensure sufficient oversight and audits; ensure the hosting organisation's financial structures are assessed prior to disbursement of funds; host funds with an professional firm
Finance	Inability of some organisations participating in the	Possible	**Moderate**	• Facilitate remote participation in meetings, and fundraise to ensure budgets enable

	coalition due to lack of financial resources			participation of all relevant members
Other	Lack of understanding of local nuances in language	Possible	**Moderate**	• Have a network of local experts who speak and understand local languages and criteria for addressing interpretation
Other	Poor communication and information sharing to and between coalition members	Possible	**Severe**	• Ensure regular and effective, structured communication channels/strategies and closing the feedback loop
Other	Over-glorification of technology and tech-solutionism	Possible	**Major**	• Tackle the problems where the people are, not to throw more technology at the problem
Other	Activities not implemented	Possible	**Severe**	• Have a coalition workplan; members to commit to implementing, fundraising, or funding the workplan activities
Other	Covid-19 pandemic restrictions	Possible	**Moderate**	• Have online meetings; reduce in-person engagements; encourage members to get vaccinations
Other	Staff turnover at secretariat	Possible	**Moderate**	• Ensure documentation of work; build capacity of coalition members; and conduct regular monitoring, evaluation, reporting and learning
Political	2022 Election	Possible	**Major**	• Monitor the political environment, and have clear short-, medium-, and long-term goals and objectives of the coalition
Political	Disagreement on direction due to poor leadership, politicisation, interference, and vested or conflict of interests of some stakeholders, e.g. on aspects such as decision-making	Possible	**Severe**	• Have strong leadership with a convenor who supports members and leverages member capacities, without competing directly with members; have an MoU with clear goals and objectives of the coalition; have documented rules on decision-making and conflict resolution; invest in leadership training; vet organisations prior to joining; invest in building trust, relationships, joint values, unity, and ethics; ensure members are champions of the coalition; have coalition structure where leadership roles are not elective,

				but purpose driven; monitor participation and influence of government regulators and those with extremist views
Political	New laws and policies	Possible	**Moderate**	• Monitor policy and legislative processes and respond appropriately
Reputational	Poorly defined and articulated purpose, goals, objectives, vision, and mission of the coalition	Unlikely	**Severe**	• Ensure clearly defined, documented, and articulated purpose, goal, objective, vision, and mission of the coalition
Safeguarding	Intimidation of minority and marginalised members by more dominant, informed, or national-level organisations	Possible	**Moderate**	• Ensuring the MoU outlines procedures for inclusive and democratic decision-making processes; ensure inclusivity, equality, and democratic participation and balanced representation from all stakeholder groups
Safeguarding	Organisations or individuals may fear associating or joining the coalition, e.g. due to fear of attacks, backlash, or pushback	Possible	**Minor**	• Declare the risks to stakeholders; have clear goals and objectives of the coalition; develop and implement safety and security plan
Safeguarding	Cyber threats to the members of the coalition's digital assets	Possible	**Moderate**	• Have in place a cybersecurity and digital resilience plan; and conduct cyber hygiene and digital security training for members
Safeguarding	Retaliation attacks for seeking and demanding accountability by government	Possible	**Moderate**	• Assess risks on a regular basis; speak based on facts/evidence and always remain objective
Stakeholder	Group dynamics, founder's syndrome, inequality of power, competition, shifting interests, and turf wars among stakeholders affect the work of the coalition leading to conflict, suspicion, and mistrust among members	Likely	**Moderate**	• Have an MoU with clear goals and objectives of the coalition; have a Code of Conduct for members; vet new members; have regular team-building, leadership training, dispute resolution procedures and capacity-building for members; build strong relationships with members; have policy on declaration of conflict of interest; avoid formalisation of the coalition; ensure transparency in communication;

				and have balanced representation and democratic decision-making
Stakeholder	Some stakeholders, e.g. platforms or government, may not be willing to meaningfully engage, take up the recommendations made, or be subjected to external oversight	Possible	**Major**	• Regularly reach out and engage officials; and have continued advocacy on the recommendations made by the coalition; clearly articulate the relationship the coalition will have with the platforms as key partners and the need for the independence of the coalition
Stakeholder	Lack of inclusivity or effective participation of key stakeholders, e.g. exclusion of marginalised or grassroots groups	Possible	**Moderate**	• Embed diversity and multi-stakeholder approaches of the coalition; ensure meaningful participation, including by being intentional; and adopt multi-stakeholder approaches to ensure inclusivity
Stakeholder	Participation of some stakeholders may adversely affect work of the coalition, e.g. inability of the coalition to criticise platforms or government publicly if they are part of the coalition	Possible	**Moderate**	• Maintain coalition independence; have working relationships with government and arm's-length engagement with the institutions. If advocacy is the mission, limit government and platform participation; if a working group approach, then include government and platforms
Stakeholder	Bureaucracy and delays in decision-making affecting coordination	Possible	**Major**	• Have a lean coalition that is agile and representative; map out all actors prior to formation of coalition; have a cap on the number of members; have clear and effective decision-making processes and procedures; and have appropriate human and financial resources in place for coordination
Stakeholder	Varying level of knowledge among stakeholders	Likely	**Major**	• Training and capacity-building to promote shared understanding of the issues
Stakeholder	Ethnic, religious, political, geo-politics, beliefs, ideological differences	Possible	**Severe**	• Objectivity and ethical approaches; members to declare interests upon joining/decision-making; leaders should be objective; have an MoU with clear goals and objectives

Notes:

* The risk type is pre-classified in the following categories: Political, Safeguarding, Stakeholder, Finance, Compliance, Reputation, Other, Covid-19.

** The risk likelihood is presented on the scale: Unlikely, Possible, Likely, and Almost certain.

*** The risk impact is presented on the scale: Minor, Moderate, Major, and Severe.

Annex B: Potential members of the coalition

Organisation	Group	Portfolio
Africa Check	Civil society	An independent fact-checking organisation working across the continent to promote accuracy of public debate and media in Africa. The organisation has published more than 1,300 fact-checked reports and fact-checked over 1,800 claims, published 180 factsheets and 47 guides on contested issues, and trained 4,500 journalists on verification best practices.
Africa Uncensored	Media	An independent media house set up by investigative journalists. It follows stories that are of importance to the country and exposes them.
Amnesty International Kenya	Civil society	An organisation dedicated to securing human rights all over the world. In Kenya, it promotes and protects civic space and civil and political rights through monitoring, research, and strategic campaigning. Our thematic areas include the right to life, human dignity, fair trial, freedom from torture, freedom of expression, and the right to peaceful assembly, association, privacy and non-discrimination.
ARTICLE 19 Eastern Africa	Civil society	ARTICLE 19 works across the region in partnership with other national and regional organisations and mechanisms to safeguard freedom of expression and information, and to create solidarity networks aimed at achieving this goal.
Bloggers Association of Kenya (BAKE)	Civil society	A community organisation that represents a group of Kenyan online content creators and seeks to empower online content creation and improve the quality of content created on the web. It also promotes online content creation and free expression in Kenya.
Centre for Human Rights and Policy Studies (CHRIPS)	Civil society	A leading international African research centre based in Kenya that conducts high-quality policy-relevant research on human rights, security, terrorism and counter-terrorism, violence, crime, and policing.
Civil Society Reference Group (CSRG)	Civil society	A membership organisation that brings together community-based organisations, national as well as international non-government organisations, and other citizen formations to advocate for the establishment of enabling legal, institutional, and operational environment for civil society organisations in Kenya.
Code for Africa/PesaCheck	Civil society	An fact-checking initiative of Code for Africa that is focused on verifying the financial and other statistical numbers quoted by public figures. PesaCheck has full-time fact-checkers in 12 countries and tracks political promises by politicians, unpacks budget and census, and builds machine learning/artificial intelligence tools to help automate verification. PesaCheck also helps watchdog media and non-government organisations to establish their own standalone fact-checking teams, and works with universities across the continent to train a new generation of civic watchdogs.
Communications Authority	Government	The regulatory authority for the communications sector in Kenya. It is responsible for facilitating the development of the information and communications sectors, including broadcasting, cybersecurity, multimedia, telecommunications, electronic commerce, and postal and

		courier services. It also licenses ICT services and enforces compliance.
Election Observers Group (ELOG)	Civil society	A long-term, permanent, and national platform, which comprises of 10 civil society organisations with the mandate of strengthening democracy in Kenya and the African region through promoting inclusive, transparent, and accountable electoral processes.
Google (Alphabet)	Private sector	Owned by Alphabet and is a multinational technology conglomerate that owns products such as Android, YouTube, Google Search, Google Suite (Gmail, Google Drive, Google Docs, Google Sheets, and Google Slides), among others.
Inter-Religious Council of Kenya (IRCK)	Faith-based	A coalition of all major faith communities in Kenya that works together to deepen inter-faith dialogue and collaboration among members for a common endeavour to mobilise the unique moral and social resources of religious people and address shared concerns.
Kenya Correspondents Association	Media	Provides a platform for media correspondents to interact, build solidarity, and enhance their profile and recognition in the media industry. The association helps its 300 correspondents to address and improve their professional and welfare needs.
Kenya Editors Guild	Media	The professional association for editors in Kenya, including senior print, broadcast and online editors, and scholars of journalism and media studies. The Guild's mission is to defend and promote media freedom and editorial independence, promote quality and ethical journalism, and provide a forum for sharing ideas and experiences that are critical in and for the media.
Kenya Human Rights Commission	Civil society	Campaigns for the entrenchment of a human rights and democratic culture in Kenya. It facilitates and supports individuals, communities, and groups to claim and defend their rights and hold state and non-state actors accountable for the protection and respect of all human rights for all people and groups.
Kenya National Commission on Human Rights (KNCHR)	Government	An independent National Human Rights Institution under the Constitution of Kenya 2010 and is the state's lead agency in the promotion and protection of human rights. The Commission acts as a watchdog over the government in the area of human rights and provides key leadership in moving the country towards a human rights state. It also investigates and provides redress for human rights violations, researches and monitors the compliance of human rights norms and standards, conducts human rights education, facilitates training, campaigns and advocates on human rights, and collaborates with other stakeholders in Kenya.
Kenya Private Sector Alliance (KEPSA)	Private sector	A membership organisation for the private sector, representing business interests across different sectors. It conducts high-level advocacy on cross-cutting law and policy-related issues, and ensures Kenya is globally competitive in doing business. It also coordinates the private sector in Kenya through various mechanisms, engages in advocacy that promotes economic growth, and conducts capacity-building of associations to strengthen, grow, and represent their sectors adequately.
KICTANet	Civil society	A multi-stakeholder think tank for people and institutions interested and involved in ICT policy and regulation. It

		conducts policy advocacy, capacity-building, research, and stakeholder engagement on ICT policy issues, while providing platforms for public to engage.
Meta Platforms Inc (Facebook Inc)	Private sector	Meta (formerly Facebook), is a multinational technology conglomerate that is the parent organisation of Facebook, Instagram, and WhatsApp, among others.
Mozilla Foundation	Civil society	Champions a healthy Internet in which privacy, openness, and inclusion are the norms, and develops trustworthy artificial intelligence through movement-building. It currently supports fellowships in Kenya that research issues such as data governance, holding artificial intelligence accountable, misinformation, and developing choice recognition for languages of underserved communities.
National Cohesion and Integration Commission (NCIC)	Government	Established to promote national identity and values, mitigate ethno-political competition and ethnically motivated violence; eliminate discrimination on ethnic, racial, and religious bases, and promote national reconciliation and healing. It promotes peace and tolerance and respect for diversity, conducts audits, capacity-building on conflict resolution, training for media, supports curriculum development for schools, conducts research, investigates complaints on hate speech and ethnic contempt, and sensitises the public on same.
National Steering Committee on Peace Building and Conflict Management	Government	An inter-agency committee within the Ministry of Interior and Coordination of National Government. Its membership comprises various state and non-state agencies working on peace and security. The Committee coordinates and consolidates efforts geared towards peacebuilding and conflict management in Kenya. The Committee implements a National Conflict Early Warning and Early Response System, capacity-building and training, holds peace forums, resolves community conflicts, coordinates peace committees and the Uwiano Platform, and conducts public sensitisation.
Office of the Attorney General	Government	The government's principal legal adviser and is responsible for the promotion of human rights and implementation of the Constitution, access to justice, good governance, anti-corruption strategies, ethics and integrity, legal education and law reform, among others. It also provides policy, coordination and oversight with regard to various legal sector institutions and promotes of the rule of law and the public interest.
Open Society Foundation (OSIEA)	Civil society	Encourages open, informed dialogue on issues of importance in Eastern Africa. Through a combination of grant-making, advocacy, and convening power, OSIEA supports and amplifies the voices of pro-democracy organisations and individuals in the region to strengthen their capacity to hold their governments accountable. This includes efforts to defend and support rights activists and pro-democracy advocates who come under attack for their work.
Strathmore University – Centre for Intellectual Property and Information Technology Law (CIPIT)	Academia	An evidence-based research and training centre based at Strathmore University, Nairobi, Kenya. Its mission is to study, create, and share knowledge on the development of intellectual property and information technology, especially as they contribute to African law and human rights.
Technology Service Providers of Kenya (TESPOK)	Private sector	A professional, non-profit organisation representing the interests of technology service providers in Kenya. It aims to influence ICT policy and regulations by engaging government

		at the relevant levels; address challenges faced by technology stakeholders and provide guidance on resolution mechanisms; provide a forum for exchange of ideas amongst industry stakeholders and development of white papers; and manage the Kenya Internet Exchange Point in line with internationally accepted best practices.
Social Media Lab (SIMElab)	Academia	A Social Media Consumption and Analytics Research Lab housed at USIU-Africa's Freida Brown Innovation Center. It offers a research and development environment.

Annex C: Interview sheet

No.	Name	Organisation	Category	Interview date
1	Agatha Ndonga	International Centre for Transitional Justice	Civil society organisation	14 October 2021
2	Allan Cheboi	Code 4 Africa	Private sector	28 October 2021
3	Annette Mbogoh	Kituo Cha Sheria	Civil society	21 October 2021
4	Anon	Media	Media organisation	25 October 2021
5	Anthony Wafula	HIVOS	Development partner	28 October 2021
6	Bernard Mugendi	Kenya Human Rights Commission (KHRC)	Civil society organisation	19 October 2021
7	Brian Kimari	CHRIPS	Civil society organisation	18 October 2021
8	Catherine Muya	ARTICLE 19 Eastern Africa	Civil society organisation	19 October 2021
9	Cheryl Akinyi	Open Society Foundation	Development partner	19 October 2021
10	Daniel Waitere	National Gender and Equality Commission	Government	20 October 2021
11	Dr. Wambui Wamunyu	US International University (USIU)	Academia	18 October 2021
12	Faith Kisinga	International Centre for Non-Profit Law (ICNL)	Civil society organisation	18 October 2021
13	Grace Bomu	Researcher	Academia	14 October 2021
14	Grace Githaiga	KICTANet	Civil society organisation	17 October 2021
15	Isaac Rutenberg	Strathmore University (CIPIT)	Academia	20 October 2021
16	James Wamathai	Bloggers Association of Kenya	Civil society organisation	21 October 2021
17	James Wanyande	National Cohesion and Integration Commission (NCIC)	Government	28 October 2021
18	John Owegi	Civil Society Reference Group (CSRG)	Civil society organisation	20 October 2021
19	Kwamchetsi Makokha	Media	Media organisation	18 October 2021
20	Lilian Kariuki	Watoto Watch Network	Civil society organisation	18 October 2021
21	Maurine Mwadime	Kenya National Commission on Human Rights (KNCHR)	Government	28 October 2021
22	Mulle Musau	Election Observation Group	Civil society organisation	19 October 2021
23	Odanga Madung’	Mozilla Foundation	Civil society organisation	18 October 2021
24	Oloo Janak	Kenya Correspondents Association (KCA)	Media organisation	17 November 2021
25	Dr. Priscah Kamungi	Consultant	Development partner	20 October 2021
26	Regina Opondo	ELOG	Civil society organisation	24 October 2021
27	Rosalia Omungo	Kenya Editors Guild (KEG)	Media organisation	18 November 2021

28	Shitemi Khamadi	Africa Uncensored	Media organisation	27 October 2021
29	Sigi Mwanzia	Digital Rights Researcher	Private sector	17 October 2021
30	Victor Ndede	Amnesty International	Civil society organisation	19 October 2021
31	William Magunga	Blogger	Media organisation	18 November 2021

Annex D: ICT status in Kenya

Table 1: ICT status in Kenya (2018–2021)

Issue/period	2018	2019	2020	2021
Population (million)	46.4	47.6	48.7	49.8[126]
Mobile subscriptions (million)	45.6	52.2	57	64.4
Mobile (SIM) penetration (%)	97.8	109.2	119.9	132.2
Internet subscriptions (million)	41.1	49.95	41.5	46.7
Internet penetration (% of population)	88.5	104.9	85.2	93.7
Broadband subscriptions (million)	20.5	22.2	22.7	27.5
Internet bandwidth (GBP)	3,278	4,655	7,393	10,218
Average speed of mobile Internet connections (Mbps)	15.1	15.53	20.64	25.06
Registered domain names	75,096	87,807	95,974	93,130

Table 2: Social media usage in Kenya (2018–2021)

Issue/period	2018	2019	2020	2021
Number of active social media users (millions)	7.7	8.2	8.8	11
Social media penetration (%) of total population	15	16	17	20.2
Number of social media users accessing via mobile phone	7.0	7.7	8.6	10.76
Social media users accessing social media via mobile (%)	90.9	93.9	98	97.8
Average number of social media accounts per Internet user	–	6.8	7.3	7.8
Average time spent on mobile Internet per day	3h 50m	–	4h 36m	4h 58m
Average time spent on social media per day	2h 54m	2h 47m	3h 23m	3h 42m

Table 3: Social media platforms users in Kenya (2018–2021)

Number of monthly active users/potential audience that can be reached using adverts (millions/year)	2018	2019	2020	2021
Facebook	7.7	7.9	8.0	9.5
YouTube	–	–	–	7.8
LinkedIn	–	2.1	2.3	2.5
Instagram	1.8	1.9	1.5	2.3
Twitter	–	0.6	0.9	1.1
SnapChat	–	0.3	0.6	1.3

Bibliography

Africa Development Bank, *Policy Brief: Minding the Gaps: Identifying Strategies to Address Gender-Based Cyber Violence in Kenya*, 2016.

Aling'o, P., 'Hate Speech And Ethnic Tension Ahead Of Kenya's 2017 Elections', Institute for Security Studies, 2016.

ANCIR, *Kenya's Keyboard Warriors*, 2021.

ARTICLE 19 Eastern Africa, Unseen Eyes Unheard Stories in Kenya, [YouTube], 2021.

ARTICLE 19 Eastern Africa, Unseen Eyes, Unheard Stories in Uganda, [YouTube], 2021.

ARTICLE 19 Eastern Africa, *Unseen Eyes, Unheard Stories: Surveillance, Data Protection, and Freedom of Expression in Kenya and Uganda during COVID-19*, 2021.

ARTICLE 19, 'Coronavirus: 75 Organisations Call on Social Media Platforms to Preserve, Publish Content Moderation Data', 2020.

ARTICLE 19, *Facebook Community Standards, Legal Analysis*, 2018.

ARTICLE 19, 'Kenya: Break the Bias to Protect Women Journalists and Human Rights Defenders', 2022.

ARTICLE 19, 'Kenya: Copyright Bill Must Respect International Standards of Free Speech', 2022.

ARTICLE 19, *Side-stepping Rights: Regulating Speech by Contract, Policy Brief*, 2018.

ARTICLE 19, *Social Media Councils: One Piece in the Puzzle of Content Moderation*, 2021.

ARTICLE 19, *Twitter Rules and Policies, Legal Analysis*, 2018.

ARTICLE 19, *Watching the Watchmen: Content Moderation, Governance, and Freedom of Expression, Policy Brief*, 2021.

ARTICLE 19, *YouTube Community Guidelines, Legal Analysis*, 2018.

Bratic, V., and Schirch, L., Why and When to Use the Media for Conflict Prevention and Peacebuilding, European Centre for Conflict Prevention, 2007.

Building Bridges Initiative (BBI), 'Unity Advisory Taskforce'.

Busolo, N. and Ngigi, S., 'Understanding Hate Speech in Kenya', *New Media and Mass Communication*, 70 (2018).

Chinmayi, C., 'Facebook's Faces', *Harvard Law Review Forum*, 135 (2022).

Code for Africa, 'Online Political Trolls', 2020.

Commission on Revenue Allocation (CRA), *Survey Report on Marginalised Areas/Counties in Kenya*, CRA Working Paper No. 2021/03, 2012.

Communications Authority of Kenya and National Cohesion and Integration Commission, *Guidelines on Prevention of Dissemination of Undesirable Bulk and Premium Rate Political Messages and Political Social Media Via Electronic Communications Networks*, 2017.

Conciliation Resources, *Pioneering Peace Pathways: Making Connections to End Violent Conflict*, Accord Issue 29, 2020.

Congressional Research Service, *Social Media: Misinformation and Content Moderation Issues for Congress*, 2021.

Council of Europe, 'Reporting on Social Media Platforms'.

David Ndii & 4 others v Attorney General & 3 others; Kenya Human Rights Commission & 2 others (Intended Amicus Curiae) [2020] eKLR.

ECPAT, INTERPOL, and UNICEF, *Disrupting Harm in Kenya: Evidence on Online Child Sexual Exploitation and Abuse*. Global Partnership to End Violence against Children, 2021.

Electronic Frontier Foundation (EFF), 'EFF, Human Rights Watch, and over 70 Civil Society Groups Ask Mark Zuckerberg to Provide all Users with Mechanism to Appeal Content Censorship on Facebook', 2018.

Elliott, C. et al., *Hate Speech, Key Concept Paper*. MeCoDEM Working Paper, White Rose University Consortium. 2016.

Equality Now, 'Kenya Just Committed to Ending Gender Based Violence in Five Years. Here's How They Plan To Do It', 2021.

European Digital Rights (EDRi), 'What You Need to Know about the Facebook Papers', 2021.

Facebook, 'Corporate Human Rights Policy'.

Freedom House, *Freedom of the Net 2018 – Kenya*, 2018.

Gavin, M., 'BBI Ruling Leaves Kenya at a Crossroads', Council on Foreign Relations, 2021.

GeoPoll. *The Reality of Fake News in Kenya*. Portland Africa, Nairobi, Kenya.

Houreld, K. and Golla, R., 'Kenyan Leaders Steer Clear of Hate Speech Before Vote, But Monitors Wary,' *Reuters*, 19 July 2017.

Human Rights Watch, 'Not Worth The Risk: Threats To Free Expression Ahead Of Kenya's 2017 Elections', 2017.

Institute for Economics & Peace, *Global Peace Index 2021: Measuring Peace In A Complex World*, 2021.

Institute for Security Studies (ISS), 'Hate Speech and Ethnic Tension Ahead of Kenya's 2017 Elections', 2016.

International Crisis Group, *Kenya's 2013 Elections: Africa Report No. 197*, 2013.

International Network Against Cyber Hate (INACH), *Monitoring Report*, 2019.

Judiciary of Kenya, *Court of Appeal Final Orders on BBI Appeals*, 2021.

Kenya Human Rights Commission (September 2008), *Violating the Vote: A report of the 2007 General Election*.

Kenya National Dialogue and Reconciliation, *Statement of Principles on Long-term Issues and Solutions*, 2008.

Kenya Transitional Justice Network, *Summary: Truth, Justice and Reconciliation Commission Report*, 2013.

KEPSA, *MKenya Daima Project Report January 2012–April 2013*.

Lotz, A., 'Profit, not free speech, governs media companies' decisions on controversy,' *The Conversation*, 2018.

Madrid-Morales, D. et al., 'Motivations for Sharing Misinformation: A Comparative Study in Six sub-Saharan African countries,' *International Journal of Communication*, 15 (2021), 1200–1219.

Makinen, M. and Kuira, M. W., 'Social Media and Post-Election Crisis in Kenya,' *Information & Communication Technology – Africa,* 13 (2008).

Merrill, J. B., 'How Facebook's Ad System Lets Companies Talk Out of Both Sides of their Mouths' *The Markup*, 2021.

Mersie, A., 'Kenya Film Board Bans Gay Documentary Calling it an Affront', *Reuters*, 27 September 2021.

Mozilla, *Inside the Shadowy World of Disinformation for Hire in Kenya*.

Mutahi, P. and Kimari, B.,'The impact of social media and digital technology on electoral violence in Kenya', *Institute of Development Studies*, 493 (2017).

Mutahi, P., and Kimari, B., 'Fake News And The 2017 Kenyan Elections', *South African Journal for Communication Theory and Research*, 46 (2020).

Muya, C., *The Law Should Work For Us Assessing Gaps in Kenya's Regulatory Framework to Build a Safer Internet for Women and Girls*, Open Internet for Democracy Initiative.

National Cohesion and Integration Commission, *A Violence Free 2022: Roadmap to Peaceful 2022 General Elections*, 2022.

National Crime Research Centre, *Gender Based Violence in Kenya*, 2014.

National Democratic Institute, *Tweets That Chill: Analyzing Online Violence Against Women In Politics*, 2019.

Newton, C., 'The Tier List: How Facebook Decides which Countries Need Protection', *The Verge*, 25 October 2021.

Njeru, A. K., Malakwen, B. and Lumala, M., 'Challenges Facing Social Media Platforms in Conflict Prevention in Kenya since 2007: A Case of Ushahidi Platform', *International Academic Journal of Social Sciences and Education*, 2 (2018).

Nyaruai, K., 'Factors Affecting Hate Speech Control For Peace Building in Kenyan Social Media. A Case of Kenyan bloggers', MA Thesis, University of Nairobi, 2015.

Odero, P. W., 'The Role of Social Media as a Tool for Peace Building and Conflict Prevention in Kenya, the Case of Nairobi County', MA Thesis, University of Nairobi, 2013.

Ogenga, F., *Social Media Literacy, Ethnicity and Peacebuilding in Kenya. Policy Brief No. 60*, Toda Peace Institute, 2019.

Orembo, L. and Gichanga, M., *Creating Safe Online Spaces For Women, Policy Brief*, Kenya ICT Action Network (KICTANet), 2020.

Owino, W., 'NCIC Rolls Out Plan to monitor Social Media Ahead of 2022 Polls', *The Standard*, 29 August 2021.

Owuor, V. and Wisor, S., *The Role of Kenya's Private Sector in Peacebuilding: The Case of the 2013 Election Cycle*, Responsibility to Protect & Business Project, One Earth Future, 2014.

Parsitau, D., 'Cyberbullying: The Digital Pandemic', *The Elephant*, 14 August 2020.

Perrigo, B., 'Inside Facebook's African Sweatshop', *Time*, 14 February 2022.

Republic of Kenya, *Building Bridges to a New Kenyan Nation*.

Republic of Kenya, *Report of the Steering Committee on the Implementation of the Building Bridges to a United Kenya Taskforce Report*, 2020.

Republic of Kenya, *The Constitution of Kenya (Amendment) Bill*, 2020.

Republic v Chief Magistrate Milimani Law Courts & 5 others Ex-parte Google Kenya Limited [2018] eKLR.

Rutenberg, I. and Sugow, A., 'Regulation of the Social Media in Electoral Democracies: A Case of Kenya', *SOAS Law Journal*, VII–I, 2020.

Siripurapu, A. and Merrow, W., 'Social Media and Online Speech: How Should Countries Regulate Tech Giants', Council on Foreign Relations, 2021.

Sissons, M., 'Our Commitment To Human Rights', Facebook, 2021.

Social Media Lab Africa (SIMElab Africa), *Social Consumption in Kenya: Trends and Practices*, US Embassy Nairobi and USIU-Africa.

Sugow, A. and Rutenberg, I., 'Safeguarding Kenya's Electoral Democracy in the Digital Age: Regulating Hate Speech and Incitement to Violence', *The Elephant*, 10 December 2021.

Tech Against Terrorism, 'The Online Regulation Series, Kenya'.

United Nations General Assembly, Thirty-fifth Session, National Report Submitted in Accordance with Paragraph 5 of the Annex to Human Rights Council Resolution 16/21: Kenya, (A/HRC/WG.6/35/KEN/1).

United Nations General Assembly, Twenty-first Session, National Report Submitted in Accordance with Paragraph 5 of the Annex to Human Rights Council Resolution 16/21: Kenya, (A/HRC/WG.6/21/KEN/1).

Van Metre, L., *Youth and Radicalization in Mombasa, Kenya: A Lexicon of Violent Extremist Language on Social Media*, PeaceTech Lab, Washington, DC.

Wahlberg, H., '"Right Now, the Lies Are ahead of us" – Maneuvering in Fake News in Kenya and Somalia,' *IMS*, 18 December 2017.

Wangui, J., 'Kenya: Why Waiguru's Case Against Internet Giant Google Stalled' *Daily Nation*, 9 June 2021.

Wood, S., 'Everything in Moderation: Artificial intelligence and Social Media Content Review,' Pillsbury Internet & Social Media Law Blog (23 March 2021).

Woolery, L., 'Companies Finally Shine A Light Into Content Moderation Practices', Centre for Democracy and Technology, 2018.

Endnotes

[1] 'Hybrid countries' are countries where elections have substantial irregularities that often prevent them from being both free and fair; government pressure on opposition parties and candidates may be common; serious weaknesses are more prevalent than in flawed democracies – in political culture, functioning of government and political participation; corruption tends to be widespread and the rule of law is weak; civil society is weak; and typically, there is harassment of and pressure on journalists, and the judiciary is not independent.

[2] See also Amnesty International, *2021 Report*.

[3] United Nations General Assembly, Thirty-fifth Session, National report submitted in accordance with paragraph 5 of the annex to Human Rights Council resolution 16/21: Kenya, (A/HRC/WG.6/35/KEN/1); and Twenty-first Session, National report submitted in accordance with paragraph 5 of the annex to Human Rights Council resolution 16/21: Kenya, (A/HRC/WG.6/21/KEN/1).

[4] They included the affronts by security agencies and aggressions experienced in the context of the Shifta war, massacres, political assassinations, detentions, torture, and ill-treatment; sexual violence; land and conflict; economic marginalisation and violation of socio-economic rights; grand corruption and economic crimes; women and children's rights violations; ethnic tensions; and violations to minority and indigenous people.

[5] The NCIC is established under the National Cohesion and Integration Act No.12 of 2008 to promote national identity and values; mitigate ethno-political competition and ethnically motivated violence; eliminate discrimination on ethnic, racial, and religious basis; and promote national reconciliation and healing.

[6] Data Reportal, Digital 2021 July Global Statshot Report, 2021; C. Mwita, *The Kenya Media Assessment 2021*, (Internews, 2021).

[7] ECPAT (End Child Prostitution in Asian Tourism) is a global network and campaign against the sexual exploitation of children.

[8] Conciliation Resources, *Pioneering Peace Pathways*, Accord Issue 29, 2020; F. Ogenga, *Social Media Literacy, Ethnicity and Peacebuilding in Kenya*, Policy Brief No. 60 (Toda Peace Institute, 2019); and P. W. Odero, 'The Role of Social Media as a Tool for Peace Building and Conflict Prevention in Kenya, the Case of Nairobi County' (University of Nairobi, 2013).

[9] For a definition of 'hate speech' under international standards on freedom of expression, see ARTICLE 19, *Hate speech explained: A toolkit*, 2019.

[10] Kenya Human Rights Commission, *Violating the Vote: A Report of the 2007 General Election*, (2008); C. Elliott et al., *Hate Speech: Key concept paper*. Working Paper (University of Leeds, 2016); M. Makinen and M. W. Kuira, Social Media and Post-Election Crisis in Kenya, (Information & Communication Technology – Africa, 2008).

[11] Lifetime prevalence is the proportion of a population that at some point in their life (up to the time of the assessment) have experienced the condition.

[12] Interview, James Wamathai, October 2021.

[13] D. Madrid-Morales et al., 'Motivations For Sharing Misinformation: A Comparative Study In Six Sub-Saharan African Countries', *International Journal of Communication*, 15 (2021).

[14] Interview, Kwamchetsi Makokha, October 2021.

[15] Interview, Oloo Janak, KCA, November 2021.

[16] Interview, Dr. Wambui Wamuyu, October 2021.

[17] Interview, Dr. Wambui Wamuyu, October 2021; Interview, Kwamchetsi Makokha, October 2021; Interview, Sigi Mwanzia, October 2021; Interview, Shitemi Khamadi, October 2021; Interview, Mulle Musau, October 2021.

[18] Interview, Allan Cheboi, October 2021; Interview, Mulle Musau, ELOG, October 2021; Interview, Catherine Muya, October 2021; Interview, Allan Cheboi, October 2021.

[19] Interview, Daniel Waitere, October 2021.

[20] Interview, Brian Kimari, October 2021.

[21] Interview, Odanga Madung, 18 October 2021.

[22] Interview, Odanga Madung, October 2021; Interview, Grace Githaiga, October 2021.

[23] Interview, Grace Bomu, researcher, October 2021.

[24] Interview, Cheryl Akinyi, October 2021.

[25] Interview, Dr. Wambui Wamuyu, 2021; Interview, Cheryl Akinyi, October 2021.

[26] Interview, Kwamchetsi Makokha, October 2021.

[27] Interview, Victor Ndede, Amnesty International, October 2021; Interview, Dr. Wambui Wamuyu, October 2021.

[28] Interview, Dr. Wambui Wamuyu, October 2021.

[29] Interview, James Wanyande, NCIC, October 2021.

[30] Interview, Maureen Mwadime, October 2021.

[31] Focus Group Discussion, Kituo cha sheria 20 October 2021.

[32] Interview, Grace Githaiga, October 2021; Interview, Bernard Mugendi, October 2021; Interview, Oloo Janak, November 2021; Interview, Daniel Waitere, October 2021; Interview, Annette Mbogoh, October 2021.

[33] Interview, Maureen Mwadime, KNCHR, October 2021; Interview, Oloo Janak, KCA, November 2021.

[34] Interview, Maureen Mwadime, KNCHR, October 2021; Interview, Oloo Janak, KCA, November 2021.

[35] Interview, Odanga Madung, Mozilla Foundation, October 2021.

[36] Interview, Kwamchetsi Makokha, October 2021.

[37] Interview, Cheryl Akinyi, October 2021.

[38] Interview, Anthony Wafula, October 2021; Interview, Isaac Rutenberg, October 2021.

[39] Interview, Sigi Mwanzia, October 2021.

[40] KICTANet, 'unseen Eyes, Unheard Stories of COVID-19 Surveillance in Kenya' [YouTube]; ARTICLE 19, 'Unseen Eyes Unheard Stories in Kenya' [YouTube].

[41] ARTICLE 19 Eastern Africa, 'Unseen Eyes Unheard Stories in Uganda' [YouTube].

[42] Google, Advertising Policies Help. Sensitive events prohibited include: 'Ads that potentially profit from or exploit a sensitive event with significant social, cultural, or political impact, such as civil emergencies, natural disasters, public health emergencies, terrorism and related activities, conflict, or mass acts of violence' and 'Ads that claim victims of a sensitive event were responsible for their own tragedy or similar instances of victim blaming; ads that claim victims of a sensitive event are not deserving of remedy or support.'

[43] Interview, Kwamchetsi Makokha, October 2021.

[44] Interview, Kwamchetsi Makokha, Media Expert; Interview, Oloo Janak, October 2021.

[45] Interview, Kwamchetsi Makokha, October 2021.

[46] Interview, Brian Kimari, October 2021; Interview, Oloo Janak, KCA, November 2021; Interview, Dr. Wambui Wamuyu, October 2021; Interview, Grace Githaiga, October 2021; Interview, Catherine Muya, October 2021; Interview, Mulle Musau, October 2021; Interview, Shitemi October 2021.

[47] Interview, Brian Kimari, October 2021.

[48] Madina Chege, Clinical officer and social activist in Kenya https://vimeo.com/388351715.

[49] Interview, Sigi Mwanzia, October 2021.

[50] Interview, Mulle Musau, October 2021; Interview, James Wamathai, October 2021; Interview, Catherine Muya, October 2021; Interview, The Magunga, October 2021; Interview, Shitemi Khamadi, October 2021.

[51] Interview, Sigi Mwanzia, October 2021.

[52] Interview, Faith Kisinga, ICNL, October 2021.

[53] Interview, Bernard Mugendi, October 2021.

[54] The activist had been arrested, detained, and charged on several occasions for his posts on social media criticising the government.

[55] Interview, Cheruyl Akinyi, October 2021.

[56] Interview, Sigi Mwanzia, October 2021.

[57] YouTube, 'NTV Kenya YouTube Channel'. Currently, the popular media station operates another YouTube channel with 1.63 million subscribers and has so far garnered 579,895,361 views. See YouTube, 'NTV Kenya'.

[58] Kenyans.co.ke, 'NTV's Channel Shut Down', 13 May 2016; T. Agumbiade, 'NTV Kenya's YouTube Channel has been terminated' 13 May 2016.

[59] The Kenya Television Network (KTN) Kenya was the first free-to-air privately owned news and entertainment TV station in the country, and is owned by the Standard Group Plc. See Standard Group, 'Our brand'; Standard Group, 'KTN News'; Standard Group, KTN Home.

[60] Kenyans.co.le, 'KTN News Twitter Account Suspended', 23 October 2018; S. Gitanga, 'KTN News Twitter Account Suspended', 23 October 2018; E. Nyambura, 'KTN News Has Twitter Account Suspended for Allegedly Buying Twitter Bots.

[61] Congressional Research Service, (January 2017), *Social Media: Misinformation and Content Moderation Issues for Congress*, 27 January 2021; Cambridge Consultants, *Use of AI in Online Content Moderation*.

[62] Interview, Victor Ndede, October 2021.

[63] Interview, Victor Ndede, October 2021.

[64] Google, 'YouTube Community Guidelines Enforcement'; Facebook, 'Taking Action'; Twitter, 'Rules Enforcement'.

[65] Republic of Kenya, National Cohesion and Integration Act, No. 12 of 2008.

[66] Republic of Kenya, Computer Misuse and Cybercrimes Act, 2018; see also the legal analysis of the Computer Misuse and Cybercrimes (Amendment) Bill by ARTICLE 19 and others.

[67] Demas Kiprono, Proposed Bill on Social Media Regulation unnecessary, *The Star*, 25 September 2019; see also ARTICLE 19, Kenya: Copyright Bill Must Respect International Standards of Free Speech, January 2022.

[68] ICT/telecommunications sector regulator.

[69] Republic of Kenya, National Cohesion and Integration Act, 2008.

[70] Republic of Kenya, Computer Misuse and Cybercrimes Act, 2018.

[71] Interview with NCIC staff, October 2021.

[72] Interview, Kwamchetsi Makokha, 18 October 2021; Interview, Isaac Rutenberg, 20 October 2021.

[73] Republic v Chief Magistrate Milimani Law Courts & 5 others Ex-parte Google Kenya Limited [2018] eKLR; Joseph Wangui, 'Kenya: Why Waiguru's Case Against Internet Giant Google Stalled' *Daily Nation*, 9 June 2021.

[74] Interview, Catherine Muya, October 2021.

[75] Interview, Sigi Mwanzia, October 2021.

[76] A. Christian, 'Facebook's Content Moderation Center in Kenya Will Create Employment', WT, 10 February 2019; K. Kangethe, 'Facebook To Hire 100 Content Reviewers For Nairobi Regional Office', *Capital FM*, 7 February 2019.

[77] Interview, Lilian Kariuki, Watoto Watch Network, 18 October 2021.

[78] Interview, James Wamathai, October 2021.

[79] Interview, Lillian Kariuki, October 2021.

[80] Interview, Grace Githaiga, October 2021.

[81] Interview, Rosalia Omungo, October 2021.

[82] Interview, Allan Cheboi, October 2021.

[83] Interview, James Wanyande, October 2021.

[84] Interview, Grace Bomu, October 2021; Interview, Faith Kisinga, October 2021.

[85] Interview, Grace Bomu, October 2021.

[86] Interview, Kwamchetsi Makokha, October 2021.

[87] Interview, Bernard Mugendi, October 2021.

[88] Exceptions include photos in the context of breastfeeding, giving birth, and after-birth moments, health-related situations (e.g. post-mastectomy, breast cancer awareness, or gender confirmation surgery) or an act of protest. Health-related is a recent addition in February 2021 after a recommendation by the Facebook Oversight Board. See N. Clegg, 'Facebook's Response to the Oversight Board's First Set of Recommendations'.

[89] Interview, Lillian Kariuki, October 2021.

[90] Interview, Grace Githaiga, 17 October 2021; Interview, Bernard Mugendi, October 2021.

[91] Interview, Shitemi Khamadi, October 2021.

[92] Interview, Catherine Muya, October 2021.

[93] Interview, Catherine Muya, October 2021.

[94] Interview, Sigi Mwanzia, October 2021; Interview, Allan Cheboi, October 2021.

[95] Interview, Victor Ndede, October 2021.

[96] Interview, Mulle Musau, October 2021.

[97] Interview, Cheryl Akinyi, October 2021.

[98] Interview, Brian Kimari, October 2021.

[99] Interview, Grace Githaiga, 17 October 2021.

[100] Interview, Kwamcheti Makokha, October 2021

[101] Interview, Dr. Wambui Wamuyu, October 2021.

[102] Interview, Sigi Mwanzia, October 2021.

[103] Interview, Sigi Mwanzia, October 2021.

[104] A classification system used by social media platform to assign particular content, e.g. words, text, images, or videos into predefined categories (e.g. racism, sexism, hate speech, non-hate speech, etc), as the basis upon which the content is moderated by human content moderators or automated moderation systems.

[105] Interview, Odanga Madung, October 2021; Interview, Kwamchetsi Makokha, October 2021.

[106] Interview, Anon, October 2021.

[107] Interview, Dr. Wambui Wamuyu, October 2021.

[108] Interview, Brian Kimari, October 2021.

[109] ARTICLE 19's views on the regulation of platforms are presented in: *Side-stepping Rights: Regulating Speech by Contract*, 2018; *Watching the Watchmen: Content Moderation, Governance and Freedom of Expression*, 2021; and Taming Big Tech, 2021. See also the Santa Clara Principles (version 2.0).

[110] National Cohesion and Integration Commission, A Violence Free 2022: Roadmap to Peaceful 2022 General Elections.

[111] Interview, Sigi Mwanzia, October 2021; Interview, Grace Bomu, 2021.

[112] Interview, Victor Ndede, Amnesty International, October 2021.

[113] Interview, Dr. Wambui Wamuyu, October 2021.

[114] Interview, Maureen Mwadime, KNCHR, 28 October 2021; Interview, Regina Opondo, October 2021.

[115] Interview, Catherine Muya, ARTICLE 19 Eastern Africa, October 2021.

[116] Interview, Sigi Mwanzia, October 2021.

[117] Interview, Regina Opondo, October 2021.

[118] Interview, Brian Kimari, October 2021.

[119] Interview, Victor Ndede, October 2021.

[120] Interview, Sigi Mwanzia, October 2021.

[121] Interview, Sigi Mwanzia, October 2021; Interview, Cheryl Akinyi, OSIEA, October 2021.

[122] Interview, James Wamathai, October 2021.

[123] Interview, Cheryl Akinyi, October 2021.

[124] Interview, Regina Opondo, October 2021.

[125] Interview, Oloo Janak, October 2021; Interview, Rosalia Omungo, November 2021.

[126] Estimated growth at 2.3%.