

Mapping of Legal Framework and Responses by Actors to Address Harmful Content Online in Kenya



AUGUST 2022

Developed by Build Up

DISCLAIMER:

©Build Up, 2022

This publication has been produced as part of UNESCO Social Media 4 Peace project funded by the European Union. 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-Share Alike 3.0 licence. You are free to copy, distribute and display this work and to make derivative works, provided you:

- 1) Credit Build Up;
- 2) Do not use this work for commercial purposes;
- 3) Distribute any works derived from this publication under a licence identical to this one. Visit this link for the full legal text of this licence: <https://creativecommons.org/licenses/by-sa/3.0/legalcode>

Acknowledgements

This report has been prepared by Build Up team comprising of Caleb Gichuhi, Kate Mytty, and Amos Atuya; with close coordination and technical review by John Okande and Adeline Hulin from UNESCO. It was copyedited and designed by Philip Emase.

Thank you to the many people who shared their experience and knowledge through interviews and workshops.

Contents

INTRODUCTION	1
ABOUT SOCIAL MEDIA 4 PEACE	4
METHODOLOGY	5
OVERVIEW OF SOCIAL MEDIA AND HARMFUL SPEECH TRENDS IN KENYA	6
OVERVIEW OF THE LEGAL FRAMEWORK FOR ADDRESSING HARMFUL SPEECH IN KENYA	9
DEFINITIONS OF HARMFUL SPEECH AND THE LIMITS TO DEFINING HARMFUL SPEECH	16
<i>Hate Speech, Disinformation, and Misinformation Signal Much Deeper-Rooted Issues in Our Societies</i>	
<i>Hate Speech, Misinformation, and Disinformation Are Interrelated</i>	
<i>Definitions of Harmful Speech within International Legislation</i>	
<i>Definitions of Harmful Speech Within Kenyan Legislation</i>	
<i>Definitions of Harmful Speech by Meta and Twitter</i>	
RESPONSES TO HARMFUL SPEECH ON SOCIAL MEDIA BY GOVERNMENT, TECHNOLOGY COMPANIES AND CIVIL SOCIETY	23
<i>Monitoring and Enforcement</i>	
<i>Monitoring and Enforcement by the Kenyan Government</i>	
<i>Monitoring and Enforcement by Social Media companies</i>	
<i>Monitoring and Interventions by Kenyan Civil society</i>	
<i>Social Media Education</i>	
<i>Peacebuilding</i>	
RECOMMENDATIONS ACROSS ACTORS	39
<i>Kenyan Government Institutions, Actions and Legislation</i>	
<i>Social Media Companies</i>	
<i>Civil Society Organizations</i>	
REFERENCES	44

Introduction

As Kenya moves towards the 2022 elections, there is an increased concern about the escalation of hate speech and misinformation, both of which have contributed to electoral violence since 1992. Although hate speech, misinformation, and disinformation have likely always existed in our societies, social media platforms have added a new dimension to the spread and access of harmful speech. This report details the legal framework / rule of law tools in place, and responses by the government, social media companies, and civil society to address illegal harmful content and potentially harmful content online in Kenya.

The evolution of media technologies has created new vehicles for the spread of hate speech and misinformation - and for the related response by government, social media companies, and civil society organizations in Kenya. Social media platforms have created a new context for hate speech, misinformation, and disinformation in Kenya.

At present, it is estimated that more than 20% of Kenya's population use social media; with the numbers increasing each year.¹ Social media, as a format, builds from other media and has unique characteristics that add specific dynamics which contribute to the spread of hate speech and misinformation - and also how harmful speech is addressed. These dynamics include:

- **Manufactured Consensus:** The use of automated or coordinated messaging from a few actors to give an illusion of the significant popularity of an idea, person, event, belief, etc. to build political support.²
- **Targeted Messaging:** This is where advertisements and messages are microtargeted and tailored to a specific group of people through demographic targeting. This may intentionally work to manipulate a group and/or be used to increase the support or cohesion of the targeted group.³
- **Algorithmic Reinforcement:** Algorithms reinforce worldviews by sharing similar content and connecting with similar people, groups, or news sources that people are already engaging with. Algorithms play to human emotion; thus, more dramatic information may spread faster and generate far greater attention than fact-based articles or information.⁴

¹ Digital in Kenya: All the Statistics You Need in 2021 – DataReportal, 1 January 2021.

² Woolley, Samuel C., and Douglas R. Guilbeault. "United States: manufacturing consensus online." *Computational propaganda: Political parties, politicians, and political manipulation on social media* (2018): 185-211.

³ Barbu, Oana. "Advertising, microtargeting and social media." *Procedia-Social and Behavioral Sciences* 163 (2014): 44-49.

⁴ Puig Larrauri, Helena, and Maude Morrison. "Understanding Digital Conflict Drivers." In *Fundamental Challenges to Global Peace and Security*, pp. 169-200. Palgrave Macmillan, Cham, 2022.

- **Significant Speed and Scale:** Social media enables people to reach a large number of social media users with few barriers. Messages sent from one context can contribute to beliefs, information and behaviours in another context.⁵

In Kenya, harmful content/speech is shared through social media platforms, contributing to the polarization of its people along ethnic, political, religious, and gender lines. This has exploited Kenya's history of unresolved land disputes; exclusion along political, gender, class and ethnic-regional lines; increased ethnicization; increased inequalities among geographies; emboldened impunity among public officials; and enhanced unequal socio-economic rights. All of these dynamics have contributed to violent conflict in the country.

It is worth noting that mechanisms developed by government, social media companies and civil society to respond to and address harmful speech, disinformation, and misinformation, are evolving. Over the past 15 years, there has been significant changes in Kenyan legislation to address harmful speech. These legislative changes have enabled the establishment of government institutions to monitor and enforce the developed legislations in Kenya. For instance, after the 2007 post-election violence in Kenya, the government, through the National Cohesion and Integration Act⁶, established the National Cohesion and Integration Commission (NCIC). One of the Commission's mandates is to eliminate all forms of ethnic or racial discrimination and to discourage the promotion of discrimination. The Act also included provisions against ethnic discrimination and harassment among other social and cohesion issues.

Even with these updates, there is a need to update Kenyan legislation and government responses to:

- a) have a clearer and more focused definition of hate speech and misinformation and to focus on dangerous speech;⁷
- b) have a clear commitment and process for protecting people's constitutional rights, especially as it relates to right to privacy; have fair and proportional consequences for people who violate Kenyan legislation;
- c) clearly address the roots and impacts of harmful speech, for instance, through beginning to integrate a restorative justice approach;
- d) address the specific dynamics enabled through social media platforms, including speed and scale, targeted messaging, algorithmic reinforcement, and manufactured consensus;
- e) require social media companies to be more transparent about their knowledge of the harms of social media platforms to users;
- f) enforce labour laws to protect social media monitors and ensure they have a healthy work environment;
- g) promote independent research on the functions and impact of social media algorithms.

Meta and Twitter are the among the most popular social media companies used in Kenya. Meta runs Facebook, Instagram and WhatsApp. Meta⁸ and Twitter⁹ have policies (on hate speech, dis/misinformation and other harmful content) that aim to reduce risk and combat the abuse of their platforms to make them safe for users. Across Kenyan legislation, United Nations policies, and social media company policies, there are no shared definitions of hate speech, misinformation, and disinformation. This allows for variance in the

⁵ Falola, Toyin. *Understanding Modern Nigeria: Ethnicity, Democracy, and Development*. Cambridge University Press, 2021.

⁶ [National and Cohesion and Integration Act 2008](#). Kenya Law. Accessed Feb 2022.

⁷ What is Dangerous Speech? Dangerous Speech Project, 2022.

⁸ Meta. [Hate speech](#). Meta, 2022.

⁹ Twitter. [Hateful conduct policy](#). Twitter, 2022.

interpretation of which policies or laws are enacted, and when and where are they enacted. The companies are making some efforts to localize their safety efforts. For instance, Meta's transparency policies have been translated into Swahili¹⁰ and Twitter has recognized Swahili as an official language within their platform.¹¹ In addition, at country level, the social media platforms (Meta¹² and Twitter¹³) now provide some information about government requests for data and advertising trends (Meta).¹⁴ There is room for social media platforms to do more to address the specific dynamics of harmful speech; this includes:

- a) being more transparent about the known harms of social media;
- b) providing more granular information about country-level trends (for instance, the proportion of harmful speech about gender);
- c) continue to expand the languages covered by the platforms; and
- d) establish clear protection policies for content moderators.

The Kenyan civil society has several actors actively addressing harmful speech on social media, and many more using text messages (SMS) as a way to support collective reporting and response. Most actors focus on monitoring social media for harmful speech; fewer efforts translate the knowledge gained through monitoring to directly respond to harmful speech. Some organizations such as the Sentinel Project are using a targeted approach with SMS to share factual messages to communities affected by misinformation in specific geographies¹⁵; other actors such as the Maskani Peacebuilders¹⁶ are working to translate in-person peacebuilding efforts onto social media. The main opportunities for Kenyan civil society are to:

- a) ensure that constitutional rights are respected, even when addressing harmful speech is the rationale for action;
- b) support the promotion of media and information literacy competencies/education about social media; and
- c) continue to deepen efforts to address the roots and impacts of harmful speech through peacebuilding and restorative justice.

¹⁰ Meta. [Transparency Center](#). Meta, 2022.

¹¹ Karimi, Faith. [Twitter Now Speaks Swahili_ Poa Sana](#). CNN, 2018.

¹² Meta Transparency Center: Government Requests for User Data: Kenya. January – December 2021. Meta, Accessed February 2022.

¹³ Twitter. Information Requests: Kenya, January – June 2021. Twitter.

¹⁴ Meta Ad Library Report – Kenya, Political or Issue Ads. Meta, 2022.

¹⁵ Tuckwood, Christopher & Saxena Raashi. [Countering rumours and misinformation with the Sentinel Project](#). Code For All, 2022

¹⁶ Owino, Sheila. Becoming a Maskani digital Peacebuilder. Medium, 2020

About Social Media 4 Peace

This publication has been produced as part of UNESCOs '[Social Media 4 Peace](#)' project funded by the European Union. This is a multi-year project focusing on [Bosnia and Herzegovina](#), Indonesia, and [Kenya](#). It aims to strengthen the resilience of civil society to address potentially harmful content spread online. In particular, Social Media 4 Peace aims to address hate speech that incites violence, while also protecting freedom of expression and enhancing the promotion of peace through digital technologies, notably social media.

Where hate speech, disinformation and misinformation have long been part of societies, the advent of social media has changed the contextual landscape. Where social media can foster the spread of misinformation, discrimination, hate speech, and polarization, social media can also be used to interrupt harm, de-escalate divisive emotions, provide additional perspectives, link distributed communities, share resources, and be a place for constructive conversations on difficult topics. The speed, scale, technical affordances, and algorithmic design of social media can contribute to the spread and escalation of hate speech and misinformation.

Where there is significant existing work to address the dynamics of social media and harmful speech, too often this conversation focuses on the global scale without the detail of the local impacts and responses within a country. The Kenya research aims to contribute to the global discussion through a nuanced understanding of what is transpiring in Kenya.

Methodology

This research draws from 13 interviews with people active in Kenya's government and civil society, along with four workshops held in Nairobi, Kisumu, and Mombasa counties. The workshops attracted the participation of 105 (54 female and 51 male) actors and stakeholders from civil society, community-based organizations, technology companies, and representatives from line ministries and agencies in government. It also draws from trends gathered from Build Up' social media listening process, along with existing research, legislation, policies, and current events.

The social media listening process focused on Twitter and Facebook, from 2017 to 2022, focusing on ethnic, political, and gender-based hateful narratives and peace messaging in the run-up to the next election cycle in August 2022. Content from the social media listening was tracked using a slur list developed from the 2013 and 2017 general elections as well as from consultations with affected communities of violent conflict and context. The literature review and interviews focused on existing legislation or company policies that address hate speech and misinformation, along with actions taken by government agencies, civil society, and technology companies to address related hate speech and misinformation narratives.

Overview of Social Media and Harmful Speech Trends in Kenya

According to the Data Report 2022, as of February 2022, Kenya had 11.75 million social media users.¹⁷ The top five social media platforms were WhatsApp at 93.5% usage by Kenyan social media users each month, Facebook at 89.6%, Instagram at 69.4%, TikTok at 60.8% and Twitter at 60.0%.¹⁸ This research focuses on Meta and Twitter to cover a significant share of users, as examples of how social media platforms are addressing harmful speech and misinformation online.

As Kenya's social media ecosystem continues to grow, it is also shaping conflict dynamics in the country. In a recent study, the NCIC - while mapping electoral violence hotspots in Kenya - identified WhatsApp, Facebook and other social media platforms as spaces where hate speech, propaganda and negative political rhetoric were rife.¹⁹ The content on social media was also cited as a significant contributor to political violence in Kenya.²⁰ While it can be argued that these trends of harmful content are a reflection of long-standing grievances of exclusion, misrepresentation, and discrimination among other societal challenges in the county, the unique characteristics that social media presents also add specific dynamics that contribute to hate speech and misinformation.

The wide usage of social media platforms impacts conflict dynamics and how harm manifests across different scales of communication - from interpersonal to public. A 2018 study in Germany for instance suggested a causal link between social media posts and physical attacks against refugees²¹. Another study focusing on the United Kingdom found "a general temporal and spatial association between online hate speech targeting race and religion and offline racially and religiously aggravated crimes"²². In Kenya, according to the Kriegler and Waki reports, some of the cases of 2007/8 post-election violence, for instance in the former Central Province, were caused by "*malicious cell phone text messages that propagated hate speech and alarming rumours*"²³.

¹⁷ Kemp, Simon. Digital in Kenya: All the Statistics You Need in 2021. Data Reportal, 15 February 2022.

¹⁸ Kemp, Simon. Digital in Kenya: All the Statistics You Need in 2021. Data Reportal, 15 February 2022.

¹⁹ National Cohesion and Integration Commission of Kenya. [Conflict Hotspot Mapping for Kenya](#). 2022.

²⁰ National Cohesion and Integration Commission of Kenya. [Conflict Hotspot Mapping for Kenya](#). 2022.

²¹ Jacobs, Josh. Does online hate drive anti-migrant violence? Financial Times. 2022

²² Williams, Matthew L., Pete Burnap, Amir Javed, Han Liu, and Sefa Ozalp. "Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime." *The British Journal of Criminology* 60, no. 1 (2020): 93-117.

²³ Konrad-Adenauer- Stiftung. [Kriegler & Waki reports](#), summarised version, revised edition. 2009

These examples point to the shift of how conflict manifests in the digital world. For instance, harmful content that was previously disseminated by a limited number of radio stations, or via text only through SMS can now be combined to include audio, video, and text, and then targeted to specific audiences via social media platforms. Information shared may be distributed from platform to platform, making it difficult to trace its origin and intention, and also enabling it to reach a wider audience on social media. Social media also creates permanent storage for information - messages sent rarely disappear; that a message lives on unless someone removes it from a social media platform. All these factors, among others, have led to the use of social media to disseminate harmful content.

There are a number of important trends in Kenyan social media. Over the last year, Build Up's social media monitoring efforts revealed that there has been an increase in hateful content on Facebook and Twitter in the lead-up to the 2022 elections in August. The increase has largely been in the form of ethnic and political-based hateful content; gender-based hate has been consistent and has not been affected by the electoral period.

Over the last year, Mozilla Fellow Odanga Madung and his colleagues released two reports on social media which uncovered how malicious, coordinated, inauthentic disinformation attacks are deployed on Twitter to silence members of civil society in Kenya.²⁴ The report's findings highlighted how disinformation influencers are paid between \$10 to \$15 per day to engage in disinformation campaigns targeting members of civil society groups.²⁵ Twitter has since taken action and suspended approximately 340 accounts after Mozilla shared their findings.^{26 27}

In Madung's third report, he shared how TikTok was being used to promote hate speech, incitement to violence, and other political disinformation.²⁸ The report identified 130 relevant TikTok videos that were viewed 4 million times.²⁹ TikTok reported having removed several videos and suspended several accounts from its platform after receiving this report.

Social media, as a vehicle for communication and disseminating information, can amplify "politics as usual". There are several examples of politicians making inflammatory remarks and encouraging ethnic balkanization.³⁰ This information is then shared on different platforms both online and offline - amplifying harmful narratives.³¹ For instance, when Kenyan politician Mithika Linturi issued an inflammatory remark in Uasin Gishu County during a political rally, it spread online.³² Mercy Corps teams in Nakuru, Naivasha and Uasin Gishu witnessed the fearful response from communities living in this region.³³ The statement was made in Uasin Gishu County, which has a significant number of people who migrate to this county for work, and have since been warned that there could be consequences if they voted for an opposing political faction.³⁴

It is now quite easy to target social media users with harmful content or with messages that reinforce a specific worldview that can increase polarization. A look at Facebook's Ad Library for Kenya in the past year shows that the majority, and in fact the top ten spenders on advertisements and targeted messaging, are politicians.³⁵ Politicians have increased their targeted messaging on social media and until other actors take up this space, it will likely continue to increase polarization.

²⁴ Madung, Odanga, with Brian Obilo. [Inside the shadowy world of disinformation for hire in Kenya](#). Mozilla Foundation, 2021.

²⁵ Madung, Odanga, with Brian Obilo. [Inside the shadowy world of disinformation for hire in Kenya](#). Mozilla Foundation, 2021.

²⁶ Madung, Odanga, with Brian Obilo. [Inside the shadowy world of disinformation for hire in Kenya](#). Mozilla Foundation, 2021.

²⁷ Madung, Odanga. [Exporting disinformation: How foreign groups peddle influence in Kenya through Twitter](#). Mozilla Foundation, 2021.

²⁸ Madung, Odanga. From Dance App to Political Mercenary: How disinformation on TikTok gaslights political tensions in Kenya. Mozilla Foundation, 2022.

²⁹ Madung, Odanga. From Dance App to Political Mercenary: How disinformation on TikTok gaslights political tensions in Kenya. Mozilla Foundation, 2022.

³⁰ Ogenga, Frederick. Interview. Conducted by Kate Mytty, February 2022.

³¹ Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

³² Apondu, Charles. Interview. Conducted by Caleb Gichuhi, February 2022.

³³ Ibid.

³⁴ Ibid.

³⁵ Meta Ad Library Report – Kenya. Meta, 2022.

Build Up's social media monitoring has uncovered efforts of coordinated messaging on Twitter from a few actors targeting political opponents to manufacture consensus and give the illusion of significant popularity of an idea that advances their political agenda. These actors deploy a hashtag early in the morning, retweeting each other for hours and by the time the larger population of social media wakes up, the hashtag is trending, giving the perception that many people are genuinely pushing and engaging with the hashtag.³⁶

Algorithmic reinforcement also plays an integral role in ways through which harmful content spreads online. A recently released Kenya conflict hotspot mapping report by the NCIC highlights the emergence of ethnic-centric social media pages within different counties that harbour harmful content.³⁷ While the emergence of homogeneous spaces online is not necessarily a problem, algorithmic reinforcement makes it easy to find in-group spaces where harm resides if one is engaging with content in their ethnic language. More dramatic information spreads faster and receives more engagement as algorithms reinforce worldviews by sharing similar content and connecting users with similar people. For instance, from Build Up's social media monitoring, harmful content, especially targeting women, has been seen to generate more attention and interaction online in Kenya.

Another unique characteristic of social media platforms is the speed and scale at which harmful content can spread. In Madung's TikTok report, 130 videos generated four million views. Build Up's monitoring efforts found out that hateful posts that could generate up to 1,500 comments a month and polarizing hashtags that would be used by 5,000 people in under an hour, reaching many more.

The social media dynamics happen within the larger Kenyan context. The digital divide in Kenya breaks down across gender, geography, and class.³⁸ There is often a language barrier. For some people in rural communities, when they see something that is written, it is assumed that the information is true simply because it is in written form. As a result, they may share that information in good faith, without realizing they are spreading misinformation.³⁹

Young people are also potentially vulnerable. It is easy for politicians to take advantage of unemployed youth by paying them to promote politicians or divisions online.⁴⁰ Beyond this, there is evidence that violent extremism and terrorism are becoming an online issue in Kenya. Social media is also being used to recruit young people into extremism and terrorism.⁴¹

Groups or individuals that are already targeted or marginalized by society are more likely to be the target of hate speech or mis/disinformation. For instance, women, LGBTQ+ people, individuals with serious diseases, or marginalized ethnicities or nationalities, are more likely to bear the brunt of hate speech and misinformation online. Because social media allows direct messages, some of this hate speech and misinformation may never be known outside of the recipients' experience. Current social media monitoring already shows that women, especially of specific ethnic groups, are being targeted in Kenya.⁴² This further opens the door to misogyny and violence towards women. For instance, when one woman stated on social media that she was sexually assaulted at a hospital, people criticized her for what she was wearing, placing the blame for the assault on the woman.⁴³

What is especially challenging about marginalized groups being targeted is that the combination of legislative limitations around discussing hate speech and misinformation, along with potential banning or silencing on social media platforms, means that targeted groups may have difficulty to honestly raise

³⁶ See Build Up's Case study of the hashtag #RutoReturnOurFood on Twitter.

³⁷ National Cohesion and Integration Commission of Kenya. Conflict Hotspot Mapping for Kenya. 2022.

³⁸ Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

³⁹ Kachwanya, Kennedy. Interview. Conducted by Caleb Gichuhi, February 2022.

⁴⁰ Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

⁴¹ Ogenga, Frederick. Interview. Conducted by Kate Mytty, February 2022.

⁴² Gichuhi, Caleb. Interview. Conducted by Kate Mytty, February 2022.

⁴³ Owino, Sheila Akinyi. Interview. Conducted by Kate Mytty, February 2022.

⁴⁴ Kiss 100 Kenya. Obinna's Instagram account being shadow banned. YouTube, 2022

awareness about their experiences without having their voices removed. In Kenya, some social media users⁴⁴ have claimed that their accounts were "shadow banned"⁴⁵ by social media companies. The companies in question have not responded to the claim. Examples in the United States however show social media users shadow banning targeted groups as a way to limit honest discussion about real-life experiences with hate speech and misinformation.⁴⁶

⁴⁵ Gadde Vijaya & Beykpour Kayvon. Setting the record straight on shadow banning. Twitter, 2018.

⁴⁶Asare, Janice Gassam. Social Media Continues To Amplify White Supremacy And Suppress Anti-Racism. Forbes, 8 January 2021.

Overview of the Legal Framework for Addressing Harmful Speech in Kenya

Over the last 15 years, the Kenyan government has responded to hate speech and misinformation on social media through new legislation and direct work by governmental institutions established through that legislation. Most legislation, especially legislation pre-dating 2013, does not explicitly address or name social media; that said, any legislation that addresses harmful speech, whether hate speech or misinformation, can also be extended to social media.

The table below highlights the main pieces of legislation that constitute the legal foundation for addressing harmful speech on social media. The various legislations have also established the NCIC, the Media Council of Kenya, the Communications Authority of Kenya, the National Computer and Cybercrimes Coordination Committee and the Independent Electoral and Boundaries Commission. Along with the Kenyan police and Kenyan courts, these government bodies are the main institutions that monitor harmful speech and enforce any responses to offences.

Each legislation applies to different constituents, as reflected in the table below. The Kenyan Constitution (2010), Penal Code (1970), the National Cohesion and Integration Act (2008), and the Computer Misuse and Cybercrime Act cover everyone in Kenya. The Communications Act focuses on messages shared through broadcasts or telecommunications. The Media Council Act focuses on journalists and media. Finally, the Election Act focuses on political parties, candidates, and leaders involved in elections.

Kenyan Legislation on Harmful Speech and Social Media

Year	Legislation	Who it applies to	Government Institutions responsible for enacting legislation
1970	Penal Code ⁴⁷	All people in Kenya.	Kenyan courts and police.
2008	National Cohesion and Integration Act ⁴⁸	All people in Kenya, explicitly including newspapers, radio stations and media enterprises.	Established the National Cohesion and Integration Commission (NCIC) as a regulatory body developed to address and reduce inter-ethnic conflict.
2009	Communications Act ⁴⁹	Anyone who sends messages through broadcasts or telecom services.	Communication Commission, to facilitate the development of the information and communications sector (including broadcasting, Multimedia, telecommunications and postal services) and electronic commerce. This was later replaced by the Communications Authority.
2010	Constitution of Kenya ⁵⁰	All people in Kenya.	Kenyan courts and police.
2011	Elections Act ⁵¹	Political parties, candidates and leaders involved in elections.	The Independent Electoral and Boundaries Commission, established under Article 88 of the Constitution.
2013	Media Council Act ⁵²	Journalists and the media.	Established the Media Council.
2013	Information and Communications (Amendment) Act ⁵³	Media and all people in Kenya.	Established the Communications Authority of Kenya, which replaced the Communications Commission (established in 2009 under the Communications Act).
2018; a petition paused enactment of legislation until 2020.	Computer Misuse and Cybercrime Act ⁵⁴	All people in Kenya.	Established the National Computer and Cybercrimes Coordination Committee.

⁴⁷Penal Code of Kenya, Revised 2012. Kenya Law, Accessed February 2022.

⁴⁸National Cohesion and Integration Act of Kenya, 2008. Kenya Law, Accessed February 2022.

⁴⁹[Kenya Communications Act](#), 2008. Kenya Law, Accessed February 2022.

⁵⁰[Constitution of Kenya, 2010](#). Kenya Law, Accessed February 2022.

⁵¹[Elections Act of Kenya](#), 2011. Independent Electoral and Boundaries Commission. Accessed February 2022.

⁵²Media Council Act, 2013. Media Council of Kenya. Accessed February 2022.

⁵³Kenya Information and Communication (Amendment), Act 2013. Kenya Law, Accessed February 2022.

⁵⁴Computer Misuse and Cybercrimes Act No. 5, 2018. Kenya Law, Accessed February 2022.

The Penal Code⁵⁵

Hate speech has been addressed by Kenyan legislation at least since the Penal Code (1970) was developed during British colonial rule; the Penal Code continues to be the main penal law that is active and has been updated as recently as 2012. The Penal Code addresses hate speech under three categories: subversive activities, libel, and defamatory matter. Under subversive activities, it is quite broad on the target and only focuses on those seen as intending to promote feelings of hatred or enmity between different races or communities in Kenya, and also hatred against a public officer. Anyone found guilty of this offence is liable to imprisonment for a term not exceeding seven years. The police and Kenyan courts are the main responsible parties for enforcing the Penal Code.

The National Cohesion and Integration Act⁵⁶

The National Cohesion and Integration Act was enacted in response to the hateful content and violence during the 2007 national elections. The NCIC Act addresses the actions of individuals, including expressions of hate speech.

The provisions in the Act against hate speech are as follows:⁵⁷ "On the offence of ethnic or racial contempt", the Act provides that:

- 1) Any person who utters words intended to incite feelings of contempt, hatred hostility, violence or discrimination against any person, group or community on the basis of ethnicity or race, commits an offence and shall be liable on conviction to a fine not exceeding one million shillings, or to imprisonment for a term not exceeding five years, or both.
- 2) A newspaper, radio station or media enterprise that publishes the utterances referred to in subsection (1) commits an offence and shall be liable on conviction to a fine not exceeding one million shillings.⁵⁸

The Act also established the National Cohesion and Integration Commission (NCIC) to "facilitate and promote equality of opportunity, good relations, harmony and peaceful coexistence between persons of the different ethnic and racial communities of Kenya, and to advise the Government on all aspects thereof."⁵⁹ Eight NCIC commissioners are nominated by and report to parliament, and appointed by the president, thus making it complicated to enforce the Act without subjectivity to being partisan.⁶⁰⁶¹

The Information and Communications (Amendment) Act⁶²

The Kenya Information and Communication (Amendment) Act 2013 updated the Kenya Information and Communications Act of 1998.

The amendment introduced hate speech and incitement restrictions and restricted freedom of the media and freedom of expression, limiting them "to the extent that the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom."⁶³ Below are some notable sub-sections in the Act:

⁵⁵ [Penal Code of Kenya](#), Revised 2012. Kenya Law, Accessed February 2022

⁵⁶ [National Cohesion and Integration Act of Kenya](#), 2008. Kenya Law, Accessed February 2022.

⁵⁷ Nderitu, Alice Wairimu. Kenya: Bridging Ethnic Divides: a Commissioner's Experience on Cohesion and Integration. Mdahalo Bridging Divides Limited, 2018.

⁵⁸ Nderitu, Alice Wairimu. Kenya: Bridging Ethnic Divides: a Commissioner's Experience on Cohesion and Integration. Mdahalo Bridging Divides Limited, 2018.

⁵⁹ Nderitu, Alice Wairimu. Kenya: Bridging Ethnic Divides: a Commissioner's Experience on Cohesion and Integration. Mdahalo Bridging Divides Limited, 2018.

⁶⁰ Nderitu, Alice Wairimu. Kenya: Bridging Ethnic Divides: a Commissioner's Experience on Cohesion and Integration. Mdahalo Bridging Divides Limited, 2018.

⁶¹ State of Internet Freedoms in Kenya. Collaboration on International ICT Policy in East and Southern Africa, 2014.

⁶² Kenya Information and Communication (Amendment), Act 2013. Kenya Law, Accessed February 2022.

⁶³ State of Internet Freedoms in Kenya. Collaboration on International ICT Policy in East and Southern Africa, 2014.

2. Subject to Article 24 of the Constitution, the right to freedom of the media and freedom of expression may be limited for the purposes, in the manner and to the extent set out in this Act and any other written law.
3. A limitation of a freedom under subsection (2) shall be limited only to the extent that the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom.
4. The right to freedom of expression shall not extend to;
 - a) the spread of propaganda for war;
 - b) incitement to violence;
 - c) the spread of hate speech; or
 - d) advocacy of hatred that -
 - i) constitutes ethnic incitement, vilification of other persons or community or incitement to cause harm; or
 - ii) is based on any ground of discrimination specified or contemplated in Article 27(4).

This Act also requires telecom service providers to register telephone subscribers and store the information, to be released either with the consent of the subscriber or for use in an investigation for criminal or civil proceedings.⁶⁴

The Communications Authority of Kenya was also created under the 2013 Act, replacing the Kenya Communications Commission from the 1998 Act.⁶⁵ The Communications Authority of Kenya is the "regulatory authority for the communications sector in Kenya", "responsible for facilitating the development of the information and communications sectors including; broadcasting, cybersecurity, multimedia, telecommunications, electronic commerce, postal and courier services."⁶⁶ Relevant to harmful speech on social media, the Communications Authority is responsible for monitoring telecommunications for hate speech, as defined within the Communications and Information (Amendment) Act. In 2020, the Communications Authority and the NCIC signed a memorandum of understanding to work together to challenge hate speech on social media, specifically asking broadcasters and bloggers not make a platform on their social media sites or broadcasts for hate speech.⁶⁷

In 2019, the Kenya Information and Communications (Amendment) Bill, 2019⁶⁸, was tabled in parliament with suggested new amendments to the 1998 Act. This new Bill was proposed with the reasons and objectives of providing for the regulation of use of social media platforms. It further seeks to introduce new sections on the licensing of social media platforms, sharing of information by licensed person(s), creates obligations for social media users, the registration of bloggers, and seeks to give responsibility to the Communications Authority to develop a bloggers code of conduct in consultation with bloggers. The Bill has not yet been enacted.

⁶⁴ State of Internet Freedoms in Kenya. Collaboration on International ICT Policy in East and Southern Africa, 2014.

⁶⁵ State of Internet Freedoms in Kenya. Collaboration on International ICT Policy in East and Southern Africa, 2014.

⁶⁶ What We Do. Communications Authority of Kenya, 2022.

⁶⁷ Communications Authority of Kenya. [Authority and National Cohesion and Integration Commission Sign Pact to Curb Hate Speech](#), 2020.

⁶⁸ Parliament of Kenya. [Kenya Information and Communications \(Amendment\) Bill 2019](#), 2019

The Constitution of Kenya⁶⁹

The Constitution of Kenya, enacted in 2010, explicitly provides several interrelated rights and fundamental freedoms, including Article 27 on equality and freedom from discrimination; Article 28 on human dignity; Article 31 on privacy; Article 32 on the freedom of conscience, religion, belief and opinion; Article 33 on the freedom of expression; Article 34 on the freedom of the media; Article 35 on the right of access to information; Article 37 on the freedom of assembly, demonstration, picketing, and petition; and Article 44 on language and culture.⁷⁰ Each of these rights are interrelated. Under freedom of expression, the Constitution explicitly forbids hate speech or advocacy of hatred.⁷¹

Article 33 (2) makes specific mention of hate speech and incitement to violence, stating that "The right to freedom of expression does not extend to-

- a) propaganda for war;
- b) incitement to violence;
- c) hate speech; or
- d) advocacy of hatred that-
 - i) constitutes ethnic incitement, vilification of others or incitement to cause harm; or
 - ii) is based on any ground of discrimination specified or contemplated in Article 27 (4)."

The grounds in Article 27 (4) include race, sex, pregnancy, marital status, health status, ethnic or social origin, colour, age, disability, religion, conscience, belief, culture, dress, language or birth.

Any efforts to address hate speech and misinformation must also be understood in the context of Kenya's 2010 Constitution.

The Media Council Act⁷²

The Media Council Act of 2013 was established to give effect to Article 34 (5)⁷³ of the Constitution of Kenya, 2010. The Act addresses hate speech under the code of conduct for the practice of journalism. The following provisions stand out:

26. Hate speech

- 1) Quoting persons making derogatory remarks based on ethnicity, race, creed, colour and sex shall not be allowed.
- 2) Racist or negative ethnic terms shall be avoided.
- 3) Careful account shall be taken on the possible effect upon the ethnic or racial group concerned, and on the population as a whole, and of the changes in public attitudes as to what is and what is not acceptable when using such terms.

The Act applies to media enterprises, journalists, media practitioners, foreign journalists, and consumers of media services (the public).⁷⁴ The Act also established and turned the responsibility for setting media standards and regulating and monitoring compliance over to the Media Council of Kenya. The Act requires that the Media Council accredits journalists - and equally, can revoke that accreditation at any time. If a

⁶⁹ Constitution of Kenya, 2010. Kenya Law, Accessed February 2022.

⁷⁰ Constitution of Kenya, 2010. Kenya Law, Accessed February 2022.

⁷¹ Constitution of Kenya, 2010. Kenya Law, Accessed February 2022.

⁷² [Media Council Act](#), 2013. Media Council of Kenya. Accessed February 2022.

⁷³ Parliament shall enact legislation that provides for the establishment of a body, which shall--

- a) be independent of control by government, political interests or commercial interests;
- b) reflect the interests of all sections of the society; and
- c) set media standards and regulate compliance with those standards.

⁷⁴ Association of Media Women in Kenya. Laws Governing Media Practice in Kenya: A Journalists' Handbook. AMWIK, 2014.

journalist does not adhere to media standards, the Media Council has the right to charge journalists with hefty fines or jail sentences. Specifically;

48. Offences

- 1) Any person that
 - a) without lawful justification fails or refuses to comply with the direction of the Council;
 - b) obstructs or hinders the Council in the exercise of its powers under this Act;
 - c) furnishes information or makes a statement to the Council which he or she knows to be false or misleading in any material particular; or
 - d) when appearing before the Council or any of its committees, for examination, makes a statement which he knows to be false or misleading in any material particular, commits an offence.
- 2) A person convicted of an offence under this section shall be liable to a fine not exceeding two hundred thousand shillings, or to imprisonment for a term not exceeding six months, or both.
- 3) A person who is convicted of an offence under subsection (2) shall be liable, for any subsequent offence, to a fine not exceeding one million shillings, or to imprisonment for a term not exceeding two years, or to both.

The Media Council Act states that a journalist is anyone who “collects, writes, edits and presents news or news articles in newspapers and magazines, radio and television broadcasts, in the internet or any other manner as may be prescribed.”

“Media” means the production of electronic and print media for circulation to the public, but does not include book publishing.⁷⁵ Thus, media could include anyone who blogs or uses social media for any wider distribution.

There are many components to the Act. In reference to hate speech and misinformation, the Act restricts the freedom of expression for journalists; freedom of expression does not extend to: propaganda for war, incitement to violence, hate speech, or advocacy of hatred that constitutes ethnic incitement, vilification of others or incitement to cause harm, or is based on any ground of discrimination.⁷⁶ This can include journalists: not quoting people who are “making derogatory remarks based on ethnicity, race, creed, colour and sex” and avoiding “racist or negative ethnic terms”. It also extends to how journalists report on ethnic, religious and sectarian violence. For instance, journalists must take time to properly verify information and also “present with due caution and restraint in a manner which is conducive to the creation of an atmosphere pleasant to national harmony, unity and peace.” This extends to images and materials also involved in violence.

The Elections Act⁷⁷

The 2011 Elections Act provides specific guidance around the structure and operations of the elections - from registration, the actual election process, resolving disputes and expectations of people involved in the elections. This includes specific guidelines around hate speech and violence, as detailed in the second schedule of the Act, under the electoral code of conduct.⁷⁸ The Act stipulates, among other requirements, that during the election cycle, people involved with the elections must publicly denounce violence and intimidation, “avoid the use of hate speech, language or any kind of action which may lead to violence or intimidation”; “refrain from any action involving violence or intimidation”, including not carrying or displaying arms or weapons; “avoid any discrimination” and “refrain from any attempt to abuse a position of power, privilege or influence” among other similar stipulations.

⁷⁵ Media Council Act, 2013. Media Council of Kenya. Accessed February 2022.

⁷⁶ Association of Media Women in Kenya. Laws Governing Media Practice in Kenya: A Journalists’ Handbook. AMWIK, 2014.

⁷⁷ Elections Act of Kenya, 2011. Independent Electoral and Boundaries Commission. Accessed February 2022.

⁷⁸ Elections Act of Kenya, 2011. Independent Electoral and Boundaries Commission. Accessed February 2022.

The Computer Misuse and Cybercrime Act⁷⁹

This Act focuses on cybercrime, establishing penalties for unauthorized interference or interception of computer systems, programs, or data; false publication of data; cyber harassment; cybersquatting; cyber terrorism; identity theft and impersonation; phishing; computer fraud; computer forgery; unauthorized disclosure of passcodes; and fraudulent use of electronic data. In addition, the Act requires service providers to aid in investigations. It prescribes significant penalties, from fines, and confiscation of relevant assets, to imprisonment when the Act is not followed.

The Bloggers Association of Kenya petitioned⁸⁰ against the Computer Misuse and Cybercrime Act on the grounds that it limited the freedoms enshrined and guaranteed by the Constitution of Kenya 2010. The Act infringes "on the freedom of opinion, freedom of expression, freedom of the media, freedom and security of the person, right to privacy, right to property and the right to a fair hearing."⁸¹ As a result of the petition, the 26 provisions of the Act were paused until 2020 when the High Court reviewed and rejected the petition.

Summary

- In general, Kenyan governmental practices and law continue to stem from colonial law, often vague, supporting surveillance of Kenyans and centring on punitive responses like fines and imprisonment. Although the legislation may address a specific instance(s) of harmful speech, the country's current legislation has few tools for addressing the roots of harmful speech. At present, within the governmental institutions, the Media Council and the NCIC are best suited to address the roots of harmful speech.
- Within each individual Act explained above, there is little clarity around how the definition of harmful speech and subsequent monitoring and enforcement also respect the freedoms enshrined in the Constitution. There is a tension between addressing harmful speech while also protecting the interrelated freedoms enshrined in Kenya's 2010 Constitution, which protects freedom of expression, freedom of media, and freedom of information.⁸² It is too easy for legislation, monitoring and enforcement to undermine rights and to increase the surveillance and policing of speech, covered by the alibi of eliminating harmful speech. The balance point is how to address harmful speech, while simultaneously protecting the fundamental promised freedoms.
- The constellation of Acts is important in that it addresses many different positions of power that may influence who and how people engage with harmful speech. At the same time, in the Media Council Act, the lack of specificity around who is part of the media can lead to greater surveillance and an unfair liability in the era of social media where anyone can create and distribute media. When everyone on social media can be identified as a part of the media, the clarity of who is responsible becomes ever more important.
- The existing legislation does not address the specific functionality of social media and how social media, as a media form, can contribute to manufactured consensus, enable targeted messaging, reinforce worldviews through algorithms, and reach a significant scale at an unprecedented speed. Existing legislation focuses more on who caused or created harmful speech versus the role of the technology in spreading hate speech and misinformation.

⁷⁹ Computer Misuse and Cybercrimes Act No. 5, 2018. Kenya Law, Accessed February 2022.

⁸⁰ Bloggers Association of Kenya. Petition against the Computer Misuse and Cybercrime Act. BAKE, 2018.

⁸¹ CFL Advocates. [Legal Update: High Court Upholds Constitutionality of Computer Misuse and Cybercrimes Act, 2018](#). Lexology, 2020.

⁸² Constitution of Kenya, 2010. Kenya Law, Accessed February 2022.

Definitions of Harmful Speech and the Limits to Defining Harmful Speech

While there is no shared definition of hate speech, misinformation, and disinformation, hate speech conceptually refers to any form of expression (e.g. speech, text, images, behaviour) that demeans or attacks or uses pejorative or discriminatory language with reference to a person or people as members of a group with shared characteristics such as ethnicity, colour, race, gender, ethnicity, religion, sexual orientation, disability or other identity factor.⁸³ Misinformation and disinformation are distinguished by one key difference. Disinformation is false information "disseminated intentionally to cause serious social harm"; misinformation is the "dissemination of false information unknowingly."⁸⁴ For brevity, harmful speech is used to encapsulate hate speech, misinformation, and disinformation; each type of speech is named when more detail is needed.

Defining hate speech, misinformation and disinformation is challenging because there is no clear line that distinguishes different types of speech. Language is always changing and often interpreted based on someone's relationship to it. Not all speech is equal - similar speech does not always result in the same impact. The existence of harmful speech, though it may cause harm, does not necessarily translate to violence. It is important that in the contextualization of harmful speech, people recognize that there is a spectrum of harmful speech, ranging from insults to incitement to violence.

When it comes to exploring the intersection of social media and harmful speech in Kenya, it is important to note that:

⁸³ United Nations, 2022. [Understanding Hate Speech](#) Accessed February 2022.

⁸⁴ Khan, Irene. Access to reliable information sources is the obvious antidote to disinformation, Special Rapporteur on freedom of opinion and expression tells Human Rights Council. United Nations Human Rights, 2021.

Hate Speech, Disinformation, and Misinformation Signal Much Deeper-Rooted Issues in Our Societies

Harmful speech cannot be divorced from the context of our societies. Understanding the trends of harmful speech is critical to exploring the potential impacts of that speech, but also to linking signals to broader societal dynamics. Personal experience, education, media, beliefs passed down from loved ones, false information, unequal power dynamics and any number of other realities in society can prime people for hate speech and misinformation - and for condoning disinformation. Where social media has contributed to the dynamics of how hate speech, disinformation and misinformation can be spread and accessed, addressing only the signals will not shift the existence of harmful speech in society. The roots of harmful speech must be tended to, alongside understanding, regulating, and addressing the specific dynamics that social media contributes to the spread of and access to harmful speech.

Hate Speech, Misinformation, and Disinformation Are Interrelated

While misinformation and hate speech are separate concepts, they are connected and are visible indicators of larger societal dynamics, including polarization. Hate speech can errantly prime a vulnerable audience to believe that a specific societal group(s) poses an existential threat, but hate speech itself is very often not the catalyst for a specific act(s) of violence.⁸⁵ Misinformation that reinforces societal stereotypes, biases, or hate speech, can become dangerous speech.⁸⁶ It can inflame conflict and increase the potential for violence and contribute to further polarization within society. Misinformation or the mischaracterization of events that did or did not occur more often provokes people to move from hating a specific group to attacking them.⁸⁷

The absence of a shared definition across Kenyan law, United Nations policy, and social media company policies is not necessarily problematic. The details in each definition create different spaces for people to thrive. For instance, in Kenyan law, sexual orientation and gender are not covered under hate speech legislation, though both are covered under Twitter⁸⁸ and Meta⁸⁹ policies. This creates a space where that speech is more likely to be protected by company policies - and thus, may allow for a much more dynamic and inclusive conversation on social media platforms.

Understanding what types of speech are covered under what definitions within Kenyan legislation is important, especially in terms of understanding which actors have or can take responsibility for addressing harmful speech, and what legislation different actors are obligated to abide by. For instance, where Meta occasionally assists the Kenyan government in legal queries, it is unclear whether or when Kenyan legislation would require Meta to assist in situations of harmful speech.

⁸⁵ Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

⁸⁶ What is Dangerous Speech? Dangerous Speech Project, 2022.

⁸⁷ Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

⁸⁸ Twitter. Hateful conduct policy. Twitter, 2022.

⁸⁹ Meta. Hate speech policy. Meta, 2022

Definitions of Harmful Speech within International Legislation

When it comes to international legislation about harmful speech, the key international commitments include: the Convention on the Prevention and Punishment of the Crime of Genocide (Genocide Convention)⁹⁰, the International Convention on the Elimination of all Forms of Racial Discrimination (ICERD)⁹¹, International Covenant on Civil and Political Rights (ICCPR)⁹², and the Rome Statute of the International Criminal Court (the Rome Statute)⁹³. Kenya has signed or is a party to all but one of these commitments - the Genocide Convention. As it applies to harmful speech, the Genocide Convention and the Rome Statute prohibit direct and public incitement to genocide. The ICERD expanded the focus on harmful speech, calling for the elimination of propaganda and organizations based on ideas or theories of the superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form.

INTERNATIONAL	META	KENYA
<ul style="list-style-type: none"> Direct and public incitement to commit genocide (Rome Statute, 1998) Advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (ICCPR, 1966) Propaganda and organisations based on ideas of theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form. (ICERD, 1965) 	<ul style="list-style-type: none"> Direct attack against people — rather than concepts or institutions— on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. 	<ul style="list-style-type: none"> Incite contempt, hatred, violence or discrimination against any person, group or community on the basis of ethnicity or race. (National Cohesion and Integration Act, 2008) Message or other matter that is grossly offensive or of an indecent, obscene or menacing character; (Communications Act, 2009) Hate speech; or (d) advocacy of hatred that—(i) constitutes ethnic incitement, vilification of others or incitement to cause harm; or (ii) is based on any ground of discrimination specified or contemplated in Article 27(4); race, sex, pregnancy, marital status, health status, ethnic or social origin, colour, age, disability, religion, conscience, belief, culture, dress, language or birth. (Kenya Constitution, 2010)

While disinformation and misinformation are not directly named in international legislation, both may fit under the existing language depending on how they are interpreted. The role social media serves in the spread of and access to harmful speech has not been addressed in UN Conventions or Statutes. What is called for by UN Conventions but not present in Kenyan law nor in the policies of social media platforms is care for people who have been affected or victimized by harmful speech and equal application of the law.

⁹⁰ United Nations. [Convention on the Prevention and Punishment of the Crime of Genocide](#). United Nations, 1948

⁹¹ United Nations Office of The High Commissioner for Human Rights. [International Convention on the Elimination of All Forms of Racial Discrimination \(ICERD\)](#). United Nations, 1965

⁹² United Nations. [International Covenant on Civil and Political Rights](#). United Nations, 1966.

⁹³ International Criminal Court. [Rome Statute of the International Criminal Court](#). International Criminal Court, 1998

Definitions of Harmful Speech Within Kenyan Legislation

While international legislation focuses on speech that incites genocide or organized and intentional propaganda, Kenya's legislation covers a range of offences. In Kenya, since the 2007 election, both misinformation and hate speech have been very broadly and vaguely defined by legislation. The key legislation in Kenya around hate speech and misinformation includes:

Year	Legislation	Disinformation	Hate Speech
Kenya Law			
1970	Penal Code		
2008	National Cohesion and Integration Act		
2009	Communications Act <i>messages through broadcasts or telecom</i>		
2010	Constitution of Kenya		
2011	Elections Act <i>Political parties, candidates and leaders involved in elections</i>		
2013	Media Council Act <i>journalist and media</i>		
2013	Communications and Information (Amendment) Act		
2018 (2020)	The Computer Misuse and Cybercrime Act		

Each Act defines harmful speech in slightly different ways and applies to different stakeholders (as listed in parentheses in the chart). Definitions of hate speech and misinformation within Kenya legislation cover a large range - from being grossly offensive (Communications Act) to advocating for hatred or sharing false information for the "purpose of causing annoyance, inconvenience or needless anxiety" (Communications Act) to "shared false information results in panic, chaos or violence" (Computer Misuse and Cybercrime Act). In April 2022, the NCIC updated its lexicon of terms they identify as hate speech⁹⁴. Immediately the lexicon was published Kenyans online started using the terms in the lexicon in their conversations that either criticized the lexicon, made fun of the terms or tried to understand the meaning of some of the terms. This resulted in an uptick in some of the banned words on social media. This shows both that not everyone agrees to the harm inherent in these terms, and also that terms covered by legislation may change, again blurring the line between what is or is not considered harmful speech.

The Kenyan Constitution (2010)⁹⁵, Penal Code (1970)⁹⁶, the National Cohesion and Integration Act (2008)⁹⁷, and the Computer Misuse and Cybercrime Act (2018)⁹⁸ cover everyone in Kenya. The Kenya Information and Communications Act⁹⁹ focuses on messages shared through broadcasts or telecommunications. The Media Council Act¹⁰⁰ focuses on journalists and media. Finally, the Election Act¹⁰¹ focuses on political parties, candidates, and leaders involved in elections. Thus, the ecosystem of legislation covers a broad - and sometimes confusing - spectrum under which hate speech and misinformation may be interpreted and for whom.

⁹⁴ National Cohesion and Integrated Commission. [Hatelex: A lexicon of hate speech terms in Kenya](#), 2022

⁹⁵ [Constitution of Kenya](#), 2010. Kenya Law, Accessed February 2022.

⁹⁶ [Penal Code](#), Kenya Law, Accessed February 2022.

⁹⁷ [National Cohesion and Integration Act 2008](#), Kenya Law, Accessed February 2022.

⁹⁸ [Computer Misuse and Cybercrimes Act, 2018](#), Kenya Law, Accessed February 2022.

⁹⁹ [Kenya Information and Communications Act 1998](#), Kenya Law, Accessed February 2022.

¹⁰⁰ [Media Council Act, 2013](#), Kenya Law, Accessed February 2022.

¹⁰¹ [Elections Act, 2011](#), Kenya Law, Accessed February 2022.

The lack of clarity and a broad spectrum of a definition of hate speech conflicts with international law. Where something may be offensive, according to international law, freedom of expression is still protected.¹⁰² The vagueness of the definitions puts more power in the arbitration of what constitutes hate speech. A critical component of the current Kenyan legislation is the role of the arbitrator(s) to determine what constitutes hate speech or misinformation; arbitrator(s) include various institutions, such as the NCIC, the Media Council, the Communications Authority and the police. Fair legislation would require independent, impartial, and informed monitoring and judicial bodies, making sure there are no biases and the law is equally applied. Examples from recent history suggest that the law is not evenly applied. There are several examples of politicians or political leaders who have knowingly skirted hate speech legislation and have not been tried under the existing law. With the current institutional structure of the Kenyan government, impartiality is difficult to establish because many of these institutions report to elected government officials.

There are several challenges with definitions in Kenyan legislation:

- a) the definitions of hate speech and misinformation are vague;
- b) there are no current requirements that monitoring institutions are independent of the political process; and
- c) the definitions do not represent the changes in technology happening through social media.

Definitions of Harmful Speech by Meta and Twitter

Meta and Twitter are among the largest social media platforms used in Kenya. Both platforms have slightly different definitions of harmful speech, which influence how they enforce what is harmful speech.

For **Meta**, the definition of hate speech focuses on direct attacks against people - "rather than concepts or institutions" - based on Meta's "protected characteristics." The protected characteristics cover different categories than are covered in the Kenyan and international legislation. Meta includes national origin, caste, sexual orientation, gender identity, and serious disease, while Kenya does not. Kenya additionally includes pregnancy, marital status, health status, social origin, colour, conscience, belief, culture, dress, language, or birth. This has implications of how Meta would monitor speech on its platforms that may conflict with Kenyan legislation.

Meta defines misinformation based on third-party fact-checkers¹⁰³ within Kenya, focusing especially on misinformation that has the potential to lead to physical harm or violence, promotes harmful health misinformation, interferes with voter or census integrity, or manipulates media. At present, the third-party fact-checkers include [Africa Check](#), [Fumbua](#), [AFP](#), and [Pesa Check](#).¹⁰⁴ Collectively they cover only English and Kiswahili in a country where more than 60 languages are spoken¹⁰⁵ showing one of the limitations of Meta's current process for identifying misinformation in the Kenyan context.

Twitter prohibits abusive behaviours,¹⁰⁶ including attempts to harass, intimidate or silence someone else's voice. Twitter also prohibits hateful conduct; users may "not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, caste, sexual orientation, gender, gender

¹⁰² Kenya Use of Hate Speech Laws. Article 19, 2020.

¹⁰³ Meta. How Facebook's Third-party Fact-Checking Program Works. Meta Blog, 1 June 2021.

¹⁰⁴ Meta. Where we have fact checking. Meta Journalism Project, 2022.

¹⁰⁵ Languages of Kenya. Wikipedia. Accessed February 2022.

¹⁰⁶ Twitter. Abusive Behavior. Twitter, 2022.

identity, religious affiliation, age, disability, or serious disease." This includes images, symbols, and profile information.¹⁰⁷

Twitter defines "misleading content ('misinformation') as claims that have been confirmed to be false by external, subject matter experts or include information that is shared in a deceptive or confusing manner."¹⁰⁸ When it comes to misinformation, there are three main categories that Twitter pays special attention to:

- 1) misleading information means that false or misleading information about COVID-19 which may lead to harm cannot be shared;¹⁰⁹
- 2) synthetic and manipulated media policy means synthetic, manipulated or out-of-context media that can deceive or confuse people and lead to harm may not be shared;
- 3) civic integrity policy does not allow the use of Twitter for manipulating or interfering with elections or other civic processes.¹¹⁰

Summary

While each of the definitions across national, international and social media policy name protected categories, there is room to further clarify what "harm of incitement" means and how it is evaluated. Resources like the Dangerous Speech Framework can create a more objective and nuanced way of identifying and categorizing speech that may increase the likelihood of harm and violence. Dangerous speech is "any form of expression that may increase the risk that its audience will condone or commit violence against members of another group."¹¹¹

Dangerousness can be analysed based on five elements: message, speaker, audience, context, and medium (the "Dangerous Speech Framework")¹¹². Under this framework, not all forms of expression hold the same weight. For instance, someone running for political office, who has a significant following, sharing a message criticizing a specific group of people may be more harmful than someone who calls someone a derogatory name in a direct message shared between two people on social media. This framework could support more nuanced legislation, policy, and programmes working to articulate and identify speech that may lower the barrier to violence.

¹⁰⁷ Twitter. Hateful Conduct Policy. Twitter, 2022.

¹⁰⁸ Twitter. COVID-19 Misleading Information Policy. Twitter, 2022.

¹⁰⁹ Twitter. COVID-19 Misleading Information Policy. Twitter, 2022.

¹¹⁰ Twitter. Civic Integrity Policy. Twitter, 2022.

¹¹¹ What is Dangerous Speech? Dangerous Speech Project, 2022.

¹¹² [Dangerous Speech: A practical Guide](#). Dangerous Speech Project, 2021

Responses to Harmful Speech on Social Media by Government, Technology Companies and Civil Society

Responses to harmful speech on social media fall into any of four categories:

- a) Monitoring for harmful speech,
- b) Enforcement of laws or policies around harmful speech,
- c) Digital literacy education, and
- d) Peacebuilding to respond to the roots causes and potential harms of harmful speech.

With any monitoring, enforcement, education and peacebuilding effort, all must be held in the delicate balance of ensuring people's rights are protected.

Monitoring and Enforcement

Monitoring and enforcement are frequent responses by different actors to identify and address harmful speech on social media. Social media arguably makes the monitoring of harmful speech more accessible than other technologies, as many voices are concentrated on singular platforms. Thus, the role of monitoring and enforcement becomes even more important when so many conversations are happening in what appears to be a public space.

Monitoring and Enforcement by the Kenyan Government

Different Kenyan governmental bodies have their own mechanisms for monitoring social media for harmful speech and enforcing relevant legislation. Each legislation has varying potential charges or responses, if someone is convicted of spreading harmful speech.

The key governmental bodies responsible for monitoring and enforcement are: National Cohesion and Integration Commission (NCIC), the Media Council of Kenya, the Communications Authority, National Computer and Cybercrimes Coordination Committee, the Independent Electoral and Boundaries Commission (IEBC), and the Kenya Police Service. Kenyan courts play an important role in determining any charges, if convicted.

Under Kenyan law, the Kenyan government has the right to seek information from social media companies to assist in law enforcement. Meta's Transparency Report shows that in 2021, the Kenyan government submitted 25 requests for information, 85% of which Meta shared data about users.¹¹³ In the first half of 2021, Twitter received one request from the Kenyan government for one account.¹¹⁴ There is no additional information on how the Kenyan government used this information or why this information was requested, or which government institution requested this information.

Below is a summary of how each Kenyan government institution monitors and enforces the relevant legislation.

Year	Offence	Potential Charges if convicted
1970	The Penal Code¹¹⁵	
	Subversive Activities (Section 77)	The following punishments may be inflicted by a court— (a) death; (b) imprisonment or, where the court so determines under the Community Service Orders Act, 1998, community service under a community service order; (c) detention under the Detention Camps Act; (e) fine; (f) forfeiture; (g) payment of compensation; (h) finding security to keep the peace and be of good behaviour; (i) any other punishment provided by this Code or by any other Act.
	Libel (Section 194)	
	Defamatory Matter (Section 195)	
2008	The National Cohesion and Integration Act¹¹⁶	
	Hate Speech (Section 13)	Any person who commits an offence under this section shall be liable to a fine not exceeding one million shillings or to imprisonment for a term not exceeding three years or to both.
	Offence of ethnic or racial contempt (Section 32)	(1) Any person who utters words intended to incite feelings of contempt, hatred, hostility, violence or discrimination against any person, group or community on the basis of ethnicity or race, commits an offence and shall be liable on conviction to a fine not exceeding one million shillings, or to imprisonment for a term not exceeding five years, or both. (2) A newspaper, radio station or media enterprise that publishes the utterances referred to in subsection (1) commits an offence and shall be liable on conviction to a fine not exceeding one million shillings.

¹¹³ Meta Transparency Center: Government Requests for User Data: Kenya. January – December 2021. Meta, Accessed February 2022.

¹¹⁴ Twitter. Information Requests: Kenya, January – June 2021. Twitter, 2022.

¹¹⁵ [Penal Code of Kenya](#), revised in 2012.

¹¹⁶ [National Cohesion and Integration Act of Kenya](#), 2008.

Year	Offence	Potential Charges if convicted
2011	The Elections Act¹¹⁷	<p>If found to have violated the Code of Conduct, a person(s) may receive:</p> <ul style="list-style-type: none"> i. a formal warning; ii. a fine determined by the Commission; iii. an order prohibiting the political party, whether permanently or for a specified period, from utilizing any public media time as have been or may be allocated to the political party for electoral purposes; iv. an order prohibiting various public engagement, including public meetings, erecting banners or placards, distributing campaign literature, etc. v. preventing people or parties from running for office, presently or in the future until fines are paid.
2013	The Media Council Act¹¹⁸	<p>Code of Conduct for Journalism</p> <ol style="list-style-type: none"> 1. If no penalty is specifically provided for, a person shall be liable, upon conviction, to a fine not exceeding two hundred thousand shillings, or to imprisonment for a term not exceeding six months, or both. 2. Where an offence under subsection (1) is committed by a body corporate, the corporate body and every principal officer or staff of the corporation directly responsible for the acts or omissions constituting the offence shall be liable to a fine not exceeding two million shillings or imprisonment for a term not exceeding two years, or to both.
2013	The Communications and Information (Amendment) Act¹¹⁹ (and the Communications Act, 2009)	<p>Freedom of the Media (Section 5B)</p> <p>If deemed responsible, the Tribunal may:</p> <ol style="list-style-type: none"> a. order the offending party to publish an apology and correction in such manner as the Tribunal may specify; b. order the return, repair, or replacement of any equipment or material confiscated or destroyed; c. make any directive and declaration on freedom of expression; d. issue a public reprimand of the journalist or media enterprise involved; e. order the offending editor of the broadcast, print or online material to publish the Tribunal's decision; f. impose a fine of not more than 20 million shillings on any respondent media enterprise and a fine of not more than 500,000 shillings on any journalist adjudged to have violated this Act; g. in its reasons for its findings, record a criticism of the conduct of the complainant in relation of the complaint, where such criticism, is in its view, warranted; h. recommend the suspension or removal from the register of the journalist involved.

¹¹⁷ [Elections Act of Kenya, 2011](#).

¹¹⁸ [Media Council Act of Kenya, 2013](#).

¹¹⁹ [Kenya Information and Communication \(Amendment\) Act 2013](#).

Year	Offence	Potential Charges if convicted
		The Computer Misuse and Cybercrime Act¹²⁰
2018	False Information (Section 22)	May be liable for a fine not exceeding 2 million shillings or to imprisonment for a term not exceeding two years, or to both.
	False Publications (Section 23)	May be liable for a fine not exceeding 5 million shillings or to imprisonment for a term not exceeding ten years, or to both.
	Cyber Harassment (Section 27)	<p>May be liable, on conviction, to a fine not exceeding 20 million shillings or to imprisonment for a term not exceeding ten years, or to both.</p> <p>A person or intermediary may also file for a restraint order against someone who is harassing them.</p> <p>The Court may order a service provider to provide any subscriber information in its possession for the purpose of identifying a person whose conduct is complained of under this section.</p> <p>A person who contravenes an order made under this section commits an offence and is liable, on conviction, to a fine not exceeding 1 million shillings or to imprisonment for a term not exceeding six months, or to both.</p>

¹²⁰[Computer Misuse and Cybercrimes Act No. 5 of 2018](#)

National Cohesion and Integration Commission (NCIC) and the Kenya Police Service

One of the most influential Kenyan government bodies regarding hate speech and misinformation is the NCIC, established through the 2008 National Cohesion and Integration Act to address and reduce inter-ethnic conflict. People found in violation of the NCIC Act or for instance, using any prohibited language identified in the NCIC's hate speech lexicon can lead to them being prosecuted under the NCIC Act or having their name added to the NCIC's public "Wall of Shame".¹²¹ Currently, NCIC has over 300 hate speech cases that are under investigation around the country.¹²² As of the publication of this report, the NCIC has 10 cases pending in courts across the country where the offence is hate speech around ethnic content, intending to close these cases soon.¹²³

It is worth noting that the NCIC works closely with the Media Council of Kenya to monitor media platforms and also with the police, who assist in conducting investigations into breaches of the NCIC Act. As part of its effort to address hate speech, the NCIC has a dedicated office for monitoring social media; they also train police in how to monitor social media, using guides like the "Police Training Manual: On the Enforcement of the Law on Hate Speech" (draft 2011).¹²⁴ With any monitoring, the government's monitoring of its citizens and residents is complicated and ripe with possibilities for overstepping the enshrined constitutional freedoms; it is too easy for the argument of combating hate speech and misinformation to be an excuse for extending surveillance. Organizations like Article 19 have expressed repeated concern over the police's monitoring of social media and implications for constitutional rights.¹²⁵

The NCIC has been criticized for not applying the law equally, by avoiding to charge political leaders and instead, targeting youth. The NCIC asserts that it does not discriminate and treats everyone spreading harmful content equally. By referencing the case of the "Pangani Six"¹²⁶ in 2016, where eight politicians were arrested and jailed in Nairobi for four nights and later released on bail for alleged hate speech¹²⁷, the NCIC has emphasized that no one is exempt from the law when it comes to using hate speech. However, despite charging the high-level officials, there has been limited additional action. For instance, two of the Pangani Six politicians were acquitted by a judge in Nairobi who ruled that the English and Kiswahili translations of the speech provided as part of the evidence in the case were incorrect.¹²⁸ This has been a common occurrence in Kenya where leaders are charged and released due to a technical issue.

The NCIC is also coordinating with other government agencies through a multi-agency approach in partnership with the Directorate of Criminal Investigations (DCI), the Ethics and Anti-Corruption Commission (EACC) and the Registrar of Political Parties (RPP). This multi-agency approach was set up to fast-track cases of hate speech and to coordinate and support the investigations of such cases. It also serves to fill the gaps that the NCIC has routinely faced in prosecuting supposed perpetrators of hate speech. If a person spews hate speech and cannot be charged under the NCIC Act because of a technicality, they can be charged through another Act as determined by a related government agency, or under the Penal Code.¹²⁹ The Penal Code has a broad definition¹³⁰ as it penalizes content that the Kenyan government believes could lead to enmity between communities in Kenya.

¹²¹ Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.

¹²² Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.

¹²³ Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.

¹²⁴ National Cohesion and Integration Commission of Kenya. Police Training Manual: On the Enforcement of the Law on Hate Speech, 2011. National Cohesion and Integration Commission of Kenya, Accessed February 2022.

¹²⁵ Article 19. Kenya Use of Hate Speech Laws. Article 19, 2020.

¹²⁶ National Cohesion and Integration Commission of Kenya. NCIC Achievements. National Cohesion and Integration Commission of Kenya, Accessed February 2022.

¹²⁷ Damary, Rita. Kalonzo wants Uhuru to drop Pangani 6 cases to show true forgiveness. The Star, 6 May 2018.

¹²⁸ Muraya, Joseph. MPs Moses Kuria And Waititu Acquitted Of Hate Speech Charges. Capital News, 20 February 2017.

¹²⁹ Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.

¹³⁰ Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.

As the NCIC continues to work with other government stakeholders to monitor harmful content in Kenya, the process from when harmful content is published to when investigations begin and a person(s) are charged continues to have gaps in time. These gaps can enable the purveyors of harmful content to evade sentencing, creating a context where some communities lose trust in the Kenyan government, and in some cases take advantage of the time gaps. These gaps in time and trust can further be exploited on social media, where people may define harmful content differently from the NCIC and one piece of harmful content can be shared by a large number of users simultaneously in the gaps of time.

The Media Council

The Media Council of Kenya monitors media outlets and journalists for their compliance with the code of conduct and professional standards set out in the Media Council Act, 2013. This includes monitoring social media. However, little information is publicly available on how the Media Council monitors social media, alongside other media channels. The Council uses the Act to define who should respect the code, this is - according to the second schedule of the Act - anyone who is a journalist, media practitioner, foreign journalist, or media enterprise. To try and remove any bias in the member selection of the Council, eight out of nine members of the Council - the Chairperson and seven members - are shortlisted and interviewed by 13 members nominated from the following organizations; the Kenya Union of Journalists, the Media Owners Association, the Kenya Editor's Guild, the Law Society of Kenya, the Kenya Correspondents Association, the Public Relations Society of Kenya, the National Gender and Equality Commission, the Association of Professional Societies in East Africa, the Consumers Federation of Kenya, the Ministry responsible for matters relating to media, the Kenya News Agency; and two persons nominated by the schools of journalism of recognized universities, one representing public universities and the other representing private universities.

To further strive for a fair and independent Council, a Media Complaints Commission as established under Section 4 of the Act is responsible for mediation or adjudication in disputes between the government and the media, the public and the media, and intra media on ethical issues; and to also ensure the adherence to high standards of journalism as provided for in the code of conduct.

The Commission chairperson is a person who holds or has held a judicial office in Kenya or who is an advocate of the High Court of Kenya of not less than ten years standing. The Commission also includes six other persons with knowledge and experience in any one of the following areas— journalism, media policy and law, media regulation, business practice and finance, the performing arts or entertainment, advertising practice; or related social sciences. A look at the Complaints Commission's website reveals that they have made 19 decisions¹³¹ on media-related complaints and disputes between the media and government, public and private enterprises from 2011 to 2021. If a journalist is found to have violated the Code of Conduct for the Practice of Journalism, their accreditation can be removed and they can be fined or imprisoned. If a media practitioner is found to have violated the Media Code of Conduct, they can be fined or imprisoned.

As such, some of the decisions made by the commission include, fining journalists and media outlets and others, including ordering the journalist to apologize to the aggrieved parties and to retract their stories.

For instance, in one of its decisions, the Commission ordered that the editor of a local newspaper retracts a story they published and offers an apology with similar prominence given to the offending story because it had been proved to be false. In addition to this, the editor was also ordered and directed to pay a fine of 50,000 shillings, or serve imprisonment for a term of three months, for making a statement to the Complaints Commission which he knew to be false or misleading. The Commission further ordered and directed the editor to pay a fine of 200,000 shillings or serve imprisonment for a term not exceeding six months, for failing to conduct himself ethically and to correct errors promptly in violation of Article 4 of the Code¹³².

¹³¹Decisions. Media Council of Kenya. Accessed February 2022

¹³²Media Council of Kenya. Complaint no. 001 of 2011, Ambassador Francis Muthaura-vs-the Standard Group and two others, 2011. Accessed March 2022

Anyone is welcome to report someone for not complying with the Media Council Act, via the [Media Council website](#). The Council also provides training to talk show hosts, journalists, online editorial content developers on handling hate speech and inflammatory language, debunking fake news and disinformation among other themes¹³³.

The Communications Authority

The Communications Authority has the ability to enforce legislation around hate speech and misinformation that is shared through different technologies - such as the internet or SMS. According to the Authority's reports, most of their monitoring data is based on reports by different telecommunications and information and communications technology (ICT) providers.¹³⁴ While there is no mention of social media monitoring on the Communications Authority website, before the 2017 election, the Authority announced that it had spent 600 million shillings to establish a social media monitoring system.¹³⁵ It is unclear whether this system still exists or whether it has been merged with the efforts by the NCIC.

If someone is deemed in violation of the Communications Act, any number of responses may happen as determined by the relevant tribunal. There may be a public apology or correction, public reprimand of the journalist or media enterprise, a declaration on freedom of expression, or potential suspension or removal of the journalist. A journalist or media enterprise may be fined up to 500,000 shillings or 20 million shillings¹³⁶, respectively.

The National Computer and Cybercrimes Coordination Committee

The National Computer and Cybercrimes Coordination Committee is explicitly involved with monitoring for cybercrimes and enforcing legislation for convicted parties. There are three relevant areas for monitoring harmful speech within the Computer Misuse and Cybercrime Act¹³⁷ : false information, false publications, and cyber harassment.

In 2021, social media monitoring for cyberbullies was expressed as a priority.¹³⁸ Even so, similar to other governmental institutions, there is little publicly available information about how the Committee is monitoring social media. Within the Act, any service providers are legally required to turn over information about anyone accused of cybercrime.

If someone is accused of false information, false publications or cyber harassment, there are various potential charges ranging from fines of two to 20 million shillings, imprisonment for two to 10 years, or a restraint filed against a perpetrator so that they do not continue to harass someone¹³⁹.

¹³³ Media council of Kenya. Media Studies 101. Accessed March 2022.

¹³⁴ Communications Authority of Kenya. Guiding Manual for the Collection of Supply-Side Data on Telecommunications/ICT 2019. Communications Authority of Kenya, Accessed February 2022.

¹³⁵ Geeska Afrika Online. Kenya to Monitor social media During Elections. Geeska Afrika Online, 2017.

¹³⁶ Kenya Information and Communications Act 1998. Kenya Law, Accessed February 2022.

¹³⁷ Computer Misuse and Cybercrimes Act No. 5, 2018. Kenya Law, Accessed February 2022.

¹³⁸ Kenya Targets Social Media "Bullies" in New Crackdown. Capital News, 4 November 2021.

¹³⁹ Computer Misuse and Cybercrimes Act No. 5, 2018. Kenya Law, Accessed February 2022

The Independent Electoral and Boundaries Commission (IEBC)

The IEBC focuses on conducting and supervising referendums and elections. Through this, they create a Code of Conduct for parties active within the election, which includes a commitment to not spread dangerous or hateful information. Thus, any party active during elections that spreads dangerous or hateful information online could be charged under the Elections Act.

There is little information on how the Commission explicitly monitors social media. That said, people can report election violations directly to the [IEBC via their website](#).

Monitoring and Enforcement by Social Media companies

Social media companies have their ways of monitoring and enforcing their policies. For identifying misinformation, **Meta** relies on third-party fact-checkers in Kenya ([Africa Check](#), [AFP Fumbua](#) and [Pesa Check](#)) to determine whether content meets Meta's misinformation threshold. Meta's main strategies for monitoring potential hate speech include having content flagged by trusted partners within Kenya, reported by users, or picked up through Meta's search function, which uses machine learning and artificial intelligence to identify "violating content".¹⁴⁰ The trusted partners are trained by Meta on its policies around hate speech, misinformation, and disinformation to inform their decisions on what content to report to Meta, in combination with the partners' Kenyan knowledge¹⁴¹. When users report specific instances of "violating content", that content is either filtered out through the same machine learning or artificial intelligence process or shared with one of Facebook's 15,000 content reviewers. Meta prioritizes what content reviewers perceive based on:

- 1) Severity:** how likely it is that the content could lead to harm, both online and offline?
- 2) Virality:** how quickly is the content being shared;
- 3) Likelihood of violating:** how likely is it that the content in question does violate Meta's policies?¹⁴²

Meta has a three-part approach to how they enforce community standards and guidelines on Instagram and Facebook:¹⁴³

- 1. Remove content that violates their policies.** This includes taking down violating content, counting strikes, restricting accounts, restricting accounts by public figures during civil unrest, disabling accounts, and removing pages and groups;
- 2. Reduce content.** For content that may not meet Meta's bar for removal, Meta may reduce the distribution of problematic content or create penalties for sharing problematic content;
- 3. Inform.** If the content is potentially sensitive or misleading, Meta may issue a warning or share additional information.¹⁴⁴

Over the last few years, Meta has been criticized over the limited efficacy of the enforcement of its policies, including its limited ability to handle the vast amount of harmful content, the amplifying role that Meta's algorithmic design serves in referring people to harmful content, and in moderators' poor working conditions.

¹⁴⁰ How technology detects violations. Meta, 19 January 2022.

¹⁴¹ Gichuhui, Caleb. Interview. Conducted by Kate Mytty, February 2022.

¹⁴² How Meta prioritizes content for review. Meta, 26 January 2022.

¹⁴³ Taking Action. Meta, 2022. Accessed February 2022.

¹⁴⁴ Transparency Center: Hate Speech. Meta, 2022. Accessed February 2022.

In 2021, Meta reported it took action on 31.5 million pieces of content globally.¹⁴⁵ Meta reports that each year, they flag more content before users flag content, suggesting that their machine learning and artificial intelligence (AI) tools are critical to their capacity and may continue to improve their capacity for identifying what Meta defines as hate speech.¹⁴⁶ While Meta identifies the challenges of identifying hate speech,¹⁴⁷ they are unclear about how their AI is trained, what languages are covered, and how AI reads imagery consistent with different cultures. In general, there is limited publicly available data on Meta's detailed approaches to harmful speech, especially at the national level.

For Kenya, Meta reports that they enacted 24 global restrictions in 2021.¹⁴⁸ It is unclear what global restrictions were enacted and on how many accounts. There is room for Kenya and Kenyans to enact policy that requires Meta to produce more granular information about what type of data is being addressed, while also respecting user privacy. For instance, Meta produced a report detailing different types of content that conflict with Austrian law.¹⁴⁹ While this type of reporting may be biased by Meta's data-collection process, this type of data would help to understand what content Meta is acting on in Kenya and how that information is being used - especially by legislative processes.

In 2021, internal Facebook research papers were released to the public highlighting the challenges of Meta's approaches to harmful content and the negative influence of their algorithmic design.¹⁵⁰ Data from Facebook's¹⁵¹ internal papers showed that Facebook takes action on "as little as 3 to 5% of hate and ~0.6% of V&I [violence and incitement] on Facebook, despite being the best in the world at it."¹⁵² Additional internal research published showed an example of how Facebook's algorithms recommended people to more polarizing content. The specific example involved a dummy account that represented "Carol" who followed a "small number of high quality/verified conservative interest follows (Fox News, Donald Trump, Melania Trump - all official pages)" and "within just one day [Facebook] Page recommendations had already devolved towards polarizing content."¹⁵³ These two points alone show the limitations of Meta's ability to identify and act on content that violates its policies and also demonstrates that Meta's algorithms can, in some circumstances, further amplify polarizing content to users who may not otherwise be exposed to that content. This research had not been publicly shared before the release of the Facebook Papers, even though Facebook had known about these limitations and potential for harm between six months and two years before the release of the Facebook Papers. The transparency around internal research on the potential limits and harms of social media platform policy enforcement and existing harms of algorithmic designs, have yet to be addressed by Kenyan legislation.

In addition, the working conditions for Meta's moderators - a critical part of Meta's monitoring and enforcement process - are increasingly under critique for harmful working conditions. Evidence from reports in Kenya and other regions show how traumatic moderating harmful content can be for employees and contractors whose work requires analysing gruesome and violent content for many hours each day.¹⁵⁴ In April 2022, a former Meta moderator in Kenya launched a lawsuit against Meta, for working conditions that undermined their constitutional rights. As of this publication, the lawsuit remains in process. It is unclear what Kenya will do to prevent future violations.

¹⁴⁵ Transparency Center: Hate Speech. Meta, 2022. Accessed February 2022.

¹⁴⁶ Meta. Transparency Center: Hate Speech. Meta, 2022. Accessed February 2022.

¹⁴⁷ Meta AI. ML Applications: AI advances to better detect hate speech. Meta, 12 May 2020.

¹⁴⁸ Meta. Transparency Center: Content Restrictions Based on Local Law: Kenya. Meta, January – July 2021 and July – December 2021. Accessed February 2022.

¹⁴⁹ Meta. Transparency Center: KoPI-G Transparency Report. Meta, October 2021.

¹⁵⁰ Dwoskin, Newmyer, and Mahtani. The Case Against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety. Washington Post, 25 October, 2021.

¹⁵¹ The internal research papers were released before Facebook was renamed to Meta.

¹⁵² Dwoskin, Newmyer, and Mahtani. The Case Against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety. Washington Post, 25 October, 2021.

¹⁵³ Dwoskin, Newmyer, and Mahtani. The Case Against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety. Washington Post, 25 October, 2021.

¹⁵⁴ Perrigo, Billy. Inside Facebook's African Sweatshop. TIME, 18 February 2022.

Twitter

When analysing a potential violation of Twitter's Rules and Policies, Twitter considers the severity of the violation and an individual's previous record of rule violations.¹⁵⁵ Twitter has a range of [enforcement options](#) and when there is a violation of hateful conduct, Twitter may respond in any number of ways, including:¹⁵⁶

- Down-ranking tweets in replies, except when the user follows the tweet author;
- Making tweets ineligible for amplification in Top search results and/or on timelines for users who do not follow the tweet author;
- Excluding tweets and/or accounts in email or in-product recommendations;
- Requiring tweet removal. For example, Twitter may ask someone to remove the violating content and serve a period in read-only mode before they can tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent suspension; and
- Suspending accounts whose primary use Twitter has determined is to engage in hateful conduct as defined in this policy, or who have shared violent threats. If the content is sensitive but does not violate Twitter's Rules and Policies, Twitter may post a warning notice, withhold a tweet based on a user's age, or withhold a tweet or account based within a country.¹⁵⁷

Twitter determines how to address misinformation based on the potential for offline harm. They take a few approaches to address misinformation including:¹⁵⁸

- **Limit amplification** or remove it from Twitter.
- **Inform and contextualize through labelling content;** prompting people to help them find additional information or consider whether they want to share information; curating Twitter moments that highlight key events, and proactively featuring informative information that serves to precheck information.
- **Providing opportunities for people to share feedback.** People in Australia, Brazil, the Philippines, South Korea, Spain, and the United States can report tweets containing misinformation. Twitter is also testing out Birdwatch in the United States, which enables an approved set of users to make contextual notes about specific tweets to provide more information. At present, this is only available to a limited number of users in the United States; the tweets and notes that users write are shared via [Birdwatch](#) on the Twitter website.

Users can submit an appeal if they believe their account or tweet was suspended in error.¹⁵⁹ Advertisers are also expected to follow Twitter's Rules and Policies.¹⁶⁰

Similar to Meta, Twitter has limited information available about trends of harmful content that are showing up on Twitter at the country scale. The global numbers are easy to find - but lack information around the detail of the content and how Twitter or national governments respond. For instance, Twitter reports a 21% increase in global compliance.¹⁶¹ To the average user, there is no way to understand how this translates to a healthier internet environment.

In addition, like Meta, Twitter also shares government information requests. In 2021, in Kenya, Twitter reports that they received one information request from the Kenyan government.¹⁶² Again, like Meta, there is no information about why that information was requested, by whom or how it was used. Thus, the information requests may be in violation with the Kenyan Constitution's right to privacy.

¹⁵⁵ Hateful Conduct Policy. Twitter, 2022

¹⁵⁶ Hateful Conduct Policy. Twitter, 2022

¹⁵⁷ Twitter. Help Center: Our Range of Enforcement Options. Twitter, 2022.

¹⁵⁸ Twitter. @TwitterMoments. Accessed February 2022.

¹⁵⁹ Twitter Help Center: About Suspended Accounts. Twitter, 2022.

¹⁶⁰ Meta Business. Ads Help Center: Hateful Content. Meta, 2022.

¹⁶¹ Twitter. Information Requests: Kenya, January – June 2021. Twitter.

¹⁶² Twitter. Information Requests: Kenya, January – June 2021. Twitter.

Twitter has a limited capacity to respond to the context and depth of Kenya's Twitter users. In 2021, Twitter had just begun to build out its continent-wide presence by recruiting a team member in Ghana.¹⁶³ This shows the limits of their capacity to respond to the entire continent, much less a single African country.

Monitoring and Interventions by Kenyan Civil society

Several well-known efforts have been implemented at various times since the 2007 election that use social media for early warning and early response. Many of these were inspired by early warning work using SMS (like Ushahidi/Uchaguzi), some have also transitioned to social media; there are fewer efforts across multiple media. At least three efforts have focused specifically on monitoring social media. From 2012 to 2013, the Umati project emerged and pioneered online hate speech monitoring in Kenya.¹⁶⁴ The programme aimed to understand hateful and inflammatory content on social media, mainly on Facebook and Twitter around the election period, not necessarily to address or counter harmful speech. This resulted in the development of the largest database of hate speech from a single country in 2013.¹⁶⁵ Sochin Agency also monitored social media posts about the Kenyan elections over six months in 2017, exploring topics around hate speech and incitement.¹⁶⁶ Also, in 2017, the PeaceTech Lab, working with Mercy Corps, established a social media monitoring mechanism and created an SMS platform to monitor for hateful language or activities; these reports were then used to inform local respondents and security agencies who could then respond locally. In addition, PeaceTech Lab disseminated peace messaging through SMS to particularly sensitive areas.¹⁶⁷

In 2021, UNESCO launched Social Media 4 Peace in Kenya to deepen understanding of the root causes, scale, and potentially harmful content and the effectiveness of various tools in addressing harmful content.¹⁶⁸ As part of this effort, Build Up is monitoring Facebook, Twitter, YouTube, TikTok and Telegram for trends in hate speech.

In addition to the aforementioned programmes that use social media directly, there are notable programmes that monitor misinformation and tensions or complaints. The Sentinel Project created Una Hakika in 2014 as an early warning system through SMS to enable citizens to share their concerns - and respond through trained responders. In advance of the 2022 election, The Sentinel Project is already working to scale up Una Hakika.¹⁶⁹ The Uwiano Platform for Peace was created by the United Nations Development Programme (UNDP) in collaboration with the NCIC, and was deployed as early as 2010 as an SMS-based early warning system for people to report tensions they were experiencing - which ranged from people not following the law, hate speech, or conflict.¹⁷⁰ In this initiative, people were also trained to be able to interpret and respond to the received requests through the platform. In 2022, Mulika Uhali fu an SMS and Mobile App based early warning and response platform is aiming to connect 30,000 trained people to respond and increase their capacity to handle more than 2,000 requests that the platform receives on matters of hate speech, violent conflict and societal tensions each day.¹⁷¹ Search for Common Ground in collaboration with UNESCO's Social Media 4 Peace project is strengthening the capacity of civil society organizations across Kenya to promote online peacebuilding narratives while equipping them with transformative digital tools and approaches for strengthening collaborative peacebuilding. The Africa Centre for People Institution and Society (ACEPIS), in partnership with UNESCO through the Social Media 4 peace project, has empowered over 1,200 young people from across 47 counties with skills and competencies on media and information literacy to tackle online hate and misinformation ahead of and after the 2022 elections in Kenya. In addition, the Media Council of Kenya has been supported by UNESCO to strengthen and increase capacities of over 200 journalists (reporters,

¹⁶³ Beykpour and Adegbite. Establishing Twitter's presence in Africa. Twitter, 12 April 2021.

¹⁶⁴ Mahihu, Morara, and Sambuli. Umati Monitoring Report: July - September 2013. iHub Research, 2013.

¹⁶⁵ Mahihu, Morara, and Sambuli. Umati Monitoring Report: July - September 2013. iHub Research, 2013.

¹⁶⁶ Sochin. Social Media Influencers, Hate Speech Monitoring, Early Warning & Response. Sochin, 2017.

¹⁶⁷ PeaceTech Lab and Mercy Corps. Election Violence Prevention. PeaceTech Lab, 2017.

¹⁶⁸ UNESCO Social Media 4 Peace: <https://en.unesco.org/news/unesco-launches-social-media-peace-project-kenya>

¹⁶⁹ The Sentinel Project. The 2022 Kenyan Election. The Sentinel Project, 2022.

¹⁷⁰ Súilíeadhbháin, Andrea Ó (editor). Leveraging Local Knowledge for Peacebuilding and Statebuilding in Africa. New York: International Peace Institute, March 2015.

¹⁷¹ Wanjohi, Grace. Interview. Conducted by Kate Mytty. February 2022.

editors, and producers) and media practitioners on conflict-sensitive reporting and data journalism for peacebuilding in Kenya. Most of these programmes explored have a limited timeline. Some were linked to specific events such as elections and had ended, or will end shortly after the event or election has passed.

Summary

When it comes to how the Kenyan government, social media companies and civil society monitor and enforce legislation, and intervene on harmful speech online, some things stand out and there are a number of changes required to be more effective in addressing harmful speech online.

- There is limited, easy to access, public information about how the Kenyan government monitors social media and how user data is stored and used. The monitoring practices may be in conflict with the Kenyan Constitution that requires the right to privacy; the lack of information about monitoring may also be in violation of the right to access to information.
- The response to any situation deemed as harmful speech focuses mostly on punitive responses - the "Wall of Shame", fines, imprisonment or suspension of journalistic activities. This has a limited impact as fines and imprisonment are unlikely to address the origin of and potential harm of harmful speech. In addition, the various punitive responses, detailed through legislation, do not identify how the threshold of what constitutes harmful speech and how this relates to the threshold of a proportionate punitive response. Thus, someone may be equally charged under the law for a minor offence, in the same proportion as someone who committed a much more grievous offence.

For instance, one element that repeatedly came up in consultative workshops is the reality of young people being hired to spread harmful speech. Because our societies are unequal, people or organizations that desire to intentionally spread harmful speech can and do take advantage of young people or people with limited economic opportunities to spread harmful speech for pay. Thus, understanding the incentive structure and origins of harmful speech is critical to also recognizing how the propagation of harmful speech fits into larger societal structures. Equally, when developing an appropriate response, it is easy to focus on who spreads hate speech, rather than the broader structures that enable people to exploit economic inequality for their own gain. There are many other examples of how a punitive response may not reflect the realities of the dynamics that lead to the origin and spread of harmful speech.

- Despite the various institutions and Acts addressing harmful speech, there is no emphasis on the protection or care of those who are victims or targets of harmful speech. The legislation focuses on the perpetrators and fails to establish inclusive redress mechanisms that provide support to those harmed by such speech. Protection and care provisions are especially important for journalists who are in the public eye and cover sensitive topics and themes, and can easily become targets of hate speech. Institutions such as the Media Council and/or the Communications Authority need to consider how these provisions can be included.
- Kenyan legislation needs to be updated to more explicitly address the dynamics of social media. This includes calling for greater transparency around the trends on social media sites, and also the growing research / knowledge known by social media companies on the true harms or limitations of social media platform designs, and the role of algorithms in amplifying harmful speech.
- By right of the Kenyan Constitution, every person has the right to equal application of the law. At present, this is not actively enforced when it comes to the application of the law to harmful speech.
- There is limited transparency - especially at the country level - around what content social media companies are seeing on their platforms, what content governments are requesting, and how that content is used. As a result, there is little useful information that can be used by civil society or government actors to guide their approaches to harmful speech thriving on social media. In addition, the lack of transparency around how people's information is being used may be in conflict with the constitutionally enshrined right to privacy. The Kenyan government and citizens could require more

transparency around the trends and use of data. For instance, if social media companies were able to provide categorical detail on trends of harmful speech, such as changes in gender-based harmful speech, civil society and government actors may be better able to address the roots of harmful speech.

- In addition, checks and balances need to be defined to limit government surveillance. Because social media companies have so much information about individual users, social media companies should be required to be more transparent and forthcoming about how user data is being utilized and especially how government entities are using social media platforms.
- Punitive enforcement is limited in its ability to address harmful speech; other strategies that concentrate on the societal root of harmful speech are necessary to change the societal dynamics around hate speech and misinformation. There is an opportunity for Kenyan government institutions and civil society organizations to explore, document and share more creative ways to address the roots and impacts of harmful speech. For instance, this can be achieved through exploring and instituting restorative justice responses to harmful speech.
- A number of civil society actors are working to address hate speech, misinformation and the root causes of each. However, many actors are siloed and disconnected - not knowing that other actors exist or competing with each other for the same resources. Historically, some actors have duplicated each other's work - and there is an opportunity to look beyond singular forms of media (like SMS only or social media only) to create a more robust ecosystem. While different approaches to the same area of work are always important, it is also helpful to create opportunities for different actors to meet, exchange information and potentially collaborate. UNESCO's Social Media 4 Peace project is already convening various actors to address harm online; similar efforts should be encouraged.
- Harmful content such as hate and dangerous speech is rooted in unaddressed community grievances, stereotypes embedded in power structures and multiple inequalities. As such, addressing them takes time. Civil Society organizations should think of approaches to embed this theme, and work in their current and future projects to make it constant and ongoing, rather than one-off projects that last six to 12 months and are linked to a brief period or event.

Social Media Education

Since the 2007 Kenyan election and its aftermath, Kenyans have become more aware of hate speech and misinformation due to significant education campaigns and related media coverage.

Within **Kenya's government**, the obvious educational arms around harmful speech and social media are the Media Council and the NCIC. The NCIC's work includes training people on how to report hate speech and misinformation to different platforms; supporting programmes to encourage youth to vote in a peaceful manner; disseminating peace messaging campaigns; and monitoring social media to identify who may be perpetuating hate speech or misinformation. Training is extended to the public and to the police, who are usually monitoring communications and making arrests. The Media Council works with journalists and different media to create ethical guidelines and training opportunities such as on media and information literacy, and conflict sensitivity among others. For instance, prior to the 2022 national election, the Media Council committed to training journalists on reporting around elections.¹⁷² This included a segment on hate speech.¹⁷³

Civil society has a strong role to play in social media education. Africa Check has been working in the Kenyan context since 2017 to reduce the circulation of false information on digital platforms and amplify accurate information. They are currently one of Facebook's trusted third-party flaggers.¹⁷⁴ By fact-checking information shared by political actors, government agencies, individuals, institutions, and other actors, they aim to encourage honest public debate and information. Their work provides a public resource for factual information and context to help people make informed decisions.¹⁷⁵ Africa Check also trains people in digital literacy and fact-checking.

Recognizing that bloggers are often accused of spreading hate speech and misinformation, the Bloggers Association of Kenya (BAKE) has established a verification and fact-checking desk to analyse bloggers' information and verify if it is true or not.¹⁷⁶ If someone posts misinformation, BAKE reaches out to them and corrects them. They also lead training on fact-checking and misinformation issues. BAKE works with bloggers so that they do not unintentionally share false information. BAKE believes it is important to educate people instead of using force so that people can self-correct harmful speech.

Beyond civil society efforts, Meta also often collaborates with civil society actors to train people on how to use Meta's platforms, and specifically on how to report hate speech and misinformation.

The education around social media platforms, or more broadly digital literacy is critical. There exist several existing resources supporting digital literacy. Education is an important component of supporting digital literacy and connecting people with tools like Africa Check. One future opportunity is to integrate peacebuilding approaches to expand people's knowledge of how they can respond in situations of harmful speech on social media.

¹⁷² Media Council of Kenya. Media Council of Kenya to Train Journalists on Elections Reporting. Media Council of Kenya, 2022.

¹⁷³ Media Council of Kenya. Media Council of Kenya to Train Journalists on Elections Reporting. Media Council of Kenya, 2022.

¹⁷⁴ Africa Check. Our Media Literacy Work. Africa Check, 2022.

¹⁷⁵ Shiundu, Alphonse (Africa Check). Interview. Conducted by Caleb Gichuhi. March 2022.

¹⁷⁶ Kachwanya, Kennedy. Interview. Conducted by Caleb Gichuhi, February 2022.

Peacebuilding

In any society, the strongest way to address harmful speech is to address the societal root causes of harmful beliefs. This is vital for building society's collective capacity to respond to situations of conflict or harm, and to be able to ensure accountability and healing. To build peace is to use non-violent means to reconcile differences and collectively transform relationships and structures in a way that is inclusive, just, and sustainable.

Sustainable peace is only possible when everyone can participate in the inclusive processes and mechanisms related to building peace. Digital technologies can contribute to harm, but can also enhance participation and inclusion in peacebuilding processes, making them more impactful. The opportunity to further integrate peacebuilding approaches within existing education, monitoring, and enforcement efforts could, over time, deepen Kenyan society's capacity for building peace. Two key approaches to learn from and integrate include:

- 1) Digital peacebuilding (or using social media within other peacebuilding efforts).
- 2) Using restorative justice to address harms caused by harmful speech.

Digital Peacebuilding

Hate speech and misinformation are signals of deeper societal dynamics and issues. Digital peacebuilding invites people to engage with these signals as an alternative to completely burying and removing them. A lot of enforcement efforts focus on the removal of hate speech or misinformation, or even individual users from social media platforms, or in some cases, a fine or imprisonment. While this may make sense for manipulated content distributed through bots or troll farms, when speaking about an individual, these tactics will result in burying the signal. This has a parallel to offline peacebuilding. It is the same dynamic in the digital world. If hate speech is taken down, it likely goes somewhere else. It may move from Facebook to Reddit, or somewhere more hidden. To build accountability and healing, people have to engage with the people who disseminate hate speech or misinformation.

Digital peacebuilding provides alternative responses to these challenges and shifts communities from relying only on the government to be the main response to hate speech and misinformation. It further provides a methodology that can be utilized across a spectrum of hate speech and misinformation, especially when content has not reached the threshold of incitement to violence. Efforts like the [Maskani Commons](#) programme launched in Western Kenya have shown viable examples of how social media can be used to promote peace and de-escalate conflict in a way that does not force people to agree but can instead foster peaceful discussions and genuine relationships.¹⁷⁷

In 2017, Kenya Elections Assistance Programme (KEAP) launched the "Zero Tolerance for Hate Speech"¹⁷⁸ campaign to discourage hate speech, hold perpetrators accountable, and generate a public commitment to non-violence through the signing of a peace pledge during the 2017 general elections. KEAP designed a digital public awareness campaign. Using a digital forensic analysis, KEAP was able to directly target the campaign to people who were more at risk or more exposed to hate speech, and engage them with nuanced messaging to create pride and unity within Kenya.

Another example is the "I Have No Tribe" blogging website launched in 2008 after its founder David Kobia, shut down another blog he was running called Mashada.com, due to uncontrollable hate speech that plagued Mashada.com, making civil discourse impossible on the platform.¹⁷⁹ "I Have No Tribe" specifically focused on constructive dialogue among Kenyans¹⁸⁰, showing posts from local and diasporic Kenyans

¹⁷⁷ Owino, Sheila Akinyi. Interview. Conducted by Kate Mytty, February 2022.

¹⁷⁸ International Foundation for Electoral Systems. Kenya Electoral Assistance Program. IFES, 2018. <https://www.ifes.org/KEAP>

¹⁷⁹ Benesch, Susan. Final Report: Countering Dangerous Speech to prevent mass violence during Kenya's 2013 elections. Dangerous Speech Project, 2014.

¹⁸⁰ Goldstein and Rotich. Digitally networked technology in Kenya's 2007-2008 post-election crisis. Berkman Center for Internet & Society at Harvard University, 2008.

discussing, debating, and unpacking the statement, "I have no tribe...I am Kenyan".

Kobia later re-opened Mashada after two months, having demonstrated that a possible response to hateful content online was to encourage constructive dialogue.¹⁸¹ In any society, those involved in and affected by conflict are best positioned to catalyse and lead change within their communities and society. These efforts exemplify how social media can be used to expand peacebuilding efforts to engage more or different people in rehumanizing relationships.

Many other efforts to promote peacebuilding have been and are being tested out in Kenya. Some have good intentions, although they could go further to impact dynamics. For instance, peace messaging that focuses on hashtag campaigns as a way to popularize peace often falls flat as it fails to create meaningful relationships. A review of the NCIC's presence on Twitter shows that the hashtag campaigns used by the Commission are mostly related to amplifying in-person events and activities as opposed to directly engaging with people online. Peace messaging, while a start, is limited; as broad peace messages often reach people who are already energized by that information and messaging. The NCIC could consider a more direct engagement strategy to address hate speech or misinformation, thus also strengthening the overall constellation of activities around the election.

Instead of hashtag campaigns, expanding digital peacebuilding programmes either through government, civil society or on social media platforms, can help people address harmful speech within their communities.

Restorative Justice

Current modes of enforcement against harmful speech are limited. Across social media platforms, Kenyan legal enforcement (and to some extent, within Kenyan civil society) typically results in someone being: blocked, banned, fined or imprisoned. None of these modes addresses the harm that has transpired, nor the roots of that harm.

There is an opportunity to expand the modes of how harmful speech on social media is addressed. Restorative justice engages people who have spread hate speech or misinformation in a dialogue to seed the possibility of accountability and healing. It creates an opportunity to identify more ways for people to both understand and address any harms that have transpired, and invites people who were negatively affected to voice their experience, and to identify what steps they need taken to address the harm.

Unlike other approaches to managing harmful speech, restorative justice asks what are the many different ways that someone may understand the harms that transpired and take responsibility? For instance, if someone published a meme that perpetuated a stereotype on social media, they may be asked to research the history of that stereotype and communicate what they learned publicly. Or someone else may be asked to participate in a circle dialogue where several people who were affected by the harmful speech can voice their experience, so that the person who published harmful speech can hear from these direct experiences and then identify a collective way forward.

For a start, restorative justice could be initiated with lighter incidents of harmful speech. Over time, as capacity and a process is built, whether it involves the Kenyan government collaborating with civil society or social media companies collaborating with civil society, or some other combination, more extreme incidents of harmful speech can be addressed with a restorative justice approach.

¹⁸¹ Zuckerman, Ethan. Kenya: Citizen media in a time of crisis. Ethan Zuckerman Blog, 20 June 2008.

Recommendations Across Actors

This section summarizes lessons learned and opportunities for addressing harmful speech by actor.

Kenyan Government Institutions, Actions and Legislation

Hate speech and misinformation legislation is important - but needs to have more specific definitions, reflect technological evolutions, and respect enshrined rights

Kenya's legal definitions of hate speech and misinformation need greater clarification and an expanded list of protected categories. Current legislation does not cover hate speech that targets people based on their gender and sexuality. Though the current legislation identifies sex as a protected characteristic, it does not include gender. In addition, by not naming sexuality but continuing to use other categorical definitions of hate speech in specific legislation means that any categorical definitions that are not included may be targeted by harmful speech. In the Constitution, freedom of expression does not extend to hate speech - thus, regardless of someone's identity, they are protected from hate speech. This conundrum must be clarified within the categories protected by Kenyan legislation.

In addition, it is unclear how people are monitored and evaluated under the vague definitions as provided in present Kenyan legislation. As a result, people are at risk of being over-surveilled, critiqued, and potentially fined or jailed for communications that may be protected under the Constitution. The further identification of a hate speech lexicon naming a specific list of words and phrases as illegal is both ineffective and makes subjective words or phrases liable to prosecution. In reality, language is fluid and people will always find coded modes of speech to skirt legal definitions of hate speech.

The Dangerous Speech Framework provides an operational framework that can support the monitoring and identification of speech that may lower the barrier to violence. Using the Dangerous Speech Framework, and the focus on the potential for inciting violence as the guiding principle behind legislation may both strengthen the likelihood of identifying harmful speech that may incite violence, thus decreasing the resources spread across monitoring and enforcement of less urgent speech.

With the emergence and growth of social media platforms, Kenyan legislation needs to continue to adapt with the evolution of the technology. The focus of current legislation is on individuals who spread harmful speech - not on the role of technological designs in amplifying harmful speech. One opportunity is to require greater transparency around the research, data, and trends from social media platforms (while also respecting individual privacy!). There is a balance between protecting people's rights to privacy, while also regulating the transparency around information and research that social media companies have regarding the potential harms that social media platforms may contribute to.

Enforcement modes to address hate speech and misinformation need to address the origins and harms of harmful speech - not just enforce punitive measures

Banning people, removing speech, fines, and imprisonment, the main vehicles for addressing harmful speech, are unlikely to address the origins of harmful speech, nor address the actual harms that transpire. Including and beginning to pilot a restorative justice practice, as a vehicle for accountability and healing, can help open up more sustainable ways of addressing the complex dynamics that lead to and result from harmful speech.

At the same time, current legislation does not distinguish between different thresholds of harm. While it is important to continue to integrate restorative justice in the short term, it is unlikely to be the only mode of response to harmful speech. Thus, enforcement measures need to require that punitive responses are proportional to the harms incurred.

Relevant to the harms incurred, there needs to be committed support for people who are targeted and/or harmed. Providing psychosocial support mechanisms may assist people in their healing process. This may also look like supporting people with security or safety measures if there is an active threat. After an incident has occurred, depending on the severity, there could be a restorative justice response where potential perpetrators can be invited to see the impacts of their actions and take responsibility for their actions. Finally, through education and training, people can be trained to be empowered to respond to incidents of harmful speech. Similar programmes such as the Maskani Commons can help build a network of people to support each other.

One of the groups that continue to be deeply and negatively affected by harmful speech on social media are social media moderators. To date in Kenya, there is not an obvious example of how labour laws protect social media moderators. This is a potential opportunity within existing labour legislation to ensure Kenyans are not harmed in their working environments. For instance, social media companies need to have clear protections for people who are hired to monitor and moderate toxic content. This includes policies on mental health. Kenya could advocate for better treatment of social media monitors.

Promote independent research on functions and the impact of social media algorithms

The connection between content shared on social media, social media algorithms and the resulting impact on beliefs, behaviours and actions continues to be explored in research. Research is partially hindered because social media platforms often provide limited data to researchers. It is also limited because of funding. Research can be supported by requiring social media platforms to provide consistent access to public information data sets on their platforms. Government agencies such as the NCIC, the Communication Authority and the National Steering Committee on Peacebuilding and Conflict Management (NSC) could invest in research that provides a Kenyan perspective on social media impacts. To fully respond to emerging threats, collaborative research on this topic needs to be ongoing as new technologies are introduced in the Kenyan media ecosystem. This can be in partnership with academic institutions, technology companies and civil society organizations. Findings from this work could inform digital literacy efforts, new legislation and amendments to existing laws.

Social Media Companies

Establish clear protection policies for content moderators

One of the realities of social media platforms is that human attention is required to monitor and moderate content being shared. Repeated and regular exposure to harmful content can affect people's health. More efforts need to be taken to support and distribute content moderation efforts. For instance, social media companies need to have clear protections for people they hire to monitor and moderate toxic content on their platforms. This includes policies on mental health and wellbeing. In instances where these moderation services are outsourced, the social media companies should ensure that the outsourced companies have transparent, inclusive and active mental health and wellness processes, and that their policies align with local labour laws. Kenya could advocate for better treatment of social media monitors.

Transparency on action taken against harmful content

Companies have begun sharing some information around the actions they are taking in addressing harmful content on their platforms. Meta's transparency centre for instance shows the volume of hate speech content they have taken action on for every three months. While this is a start, more needs to be done. More information outlining the topic or theme of the harmful content taken down can be shared, among government and trusted civil society organizations or individuals, to help a more localized understanding and responses to toxic content. For instance, Facebook could share what percentage of harmful posts taken down were focused on gender and ethnicity or religious intolerance. It might also be useful to show, at country-level, content that was flagged by the company and content that was reported by users. This type of information can be used to inform programmes on media literacy and understand if /how users are applying skills around reporting, blocking or directly engaging with harmful content.

Transparency about known harms

Social media companies have a duty to protect its users and society. They are regularly seeing and researching the impacts of social media platforms; for instance, the information that was leaked through the Facebook Papers shows the extent of Meta's knowledge about the realities of their algorithms. As a result, social media companies need to unveil and share their knowledge about the harm caused through their algorithmic design to prevent further damage and inform their peer social media companies so that they can equally respond to similar harms.

Continue to expand languages covered

Social media should continue to provide resources and guidelines in relevant languages to users. With this information available in users' primary language, it will help users to fully understand the different functionalities of social media platforms that they can use to respond to harmful content.

For instance, right now Meta only recognizes Swahili and English as the main languages for Kenya. This does not include Somali for instance, despite the large community of Somali people in Kenya who have been targeted through social media by violent extremist groups. By expanding the languages covered to include Somali, Meta could support Somalis in accessing the necessary information.

Civil Society Organizations

Build capacity for collective response to harmful incidents and the roots of harmful speech

This includes deepening and expanding efforts around digital literacy, digital peacebuilding, and restorative justice. Digital literacy programmes are important and could be expanded with more training and practice on digital peacebuilding. The combination of digital literacy and digital peacebuilding can provide people with more options for how they can respond to harmful speech on social media. In addition, Kenyan civil society is well suited to pilot and model a restorative justice approach to incidents of harmful content. By piloting different approaches for restorative justice, Kenyan civil society may influence the government to adapt similar approaches.

Share digital literacy knowledge more widely and regularly

Social media and communications are always changing. Regular digital literacy and sensitization programmes can help people continue to be informed and empowered to engage thoughtfully on social media without manipulation. Social media can be a resource to educate the public quickly and widely about digital literacy. For instance, educating people about deep fakes, which are manipulated videos made to look real. As technology has changed, deep fakes are increasingly more accessible for a wider variety of people to use in any number of ways - and sometimes used to manipulate people and spread false information. Social media campaigns about deep fakes may create a wider awareness of the potential ways people can be manipulated. Other digital peacebuilding efforts could include training on how people can recognize and respond within their communities to hate speech and misinformation, and the steps social media users can take to report harmful information through social media platforms.

Build out additional social media and harmful speech training across more languages

In Kenya, a lot of the training and information about social media is predominantly available through English and Kiswahili.¹⁸² To support greater access to social media training, this training needs to be available across more languages and in ways that are accessible even for people with more limited access to the internet.

The below recommendations cut across actors, including civil society and the Kenyan government.

Address the specific dynamics enabled through social media platforms, including speed and scale, targeted messaging, algorithmic reinforcement and manufactured consensus

This requires a multi-stakeholder approach. The Kenyan government is best suited to create legislation that reflects the speed and scale and complexity of social media platforms. For instance, the government can require a faster take-down time for dangerous content, and require greater transparency about algorithmic design, to understand and legislate against negative impacts. In addition, they can require more transparency to how targeted messaging is being used.

Social media companies are best suited to research the effects of their algorithms and create measures to prevent algorithmic reinforcement that amplifies harmful content. They can also help users identify what content is manufactured so users can make informed decisions. For instance, social media companies could highlight content that manufactures consensus so that users know what is coordinated versus more organic content.

Civil society can use targeted messaging to advance their digital peacebuilding efforts. In addition, all three actors can support education about social media advances, for instance, introducing users more directly to the realities of social media dynamics.

Continue and diversify engagement around hate speech and misinformation even after the Kenyan elections, while still upholding the Constitution

Future work on harmful speech will require collective cooperation between Kenyan government, civil society and social media companies. No single actor is independently able to enforce constitutional rights; as the social media technologies continue to diversify, it is the responsibility of each of these actors to uphold the Constitution and to ensure that their counterparts do the same. Each actor has a different capacity to check and balance the other actors on how they uphold the Constitution.

In addition, a lot of the work in Kenya by the government and civil society that seeks to address hate speech and misinformation is increased and elevated around the national elections and often suspended when elections conclude. Where escalated conflict is possible during elections and evident given the track record around recent elections, this limited engagement is not enough. A lot of the election-based efforts focus on monitoring and research around trends in hate speech and misinformation. This is helpful in understanding what is happening - but more direct engagement needs to be seen on a more extended basis than around the elections.

An essential part of extending this work is providing funding to do so. After elections, funding for work addressing hate speech, misinformation and related contexts on social media tends to shrink, leaving many potential actors without resources to support their work.¹⁸³

¹⁸² Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.

¹⁸³ Interview with Charles Apondu, Mercy Corps

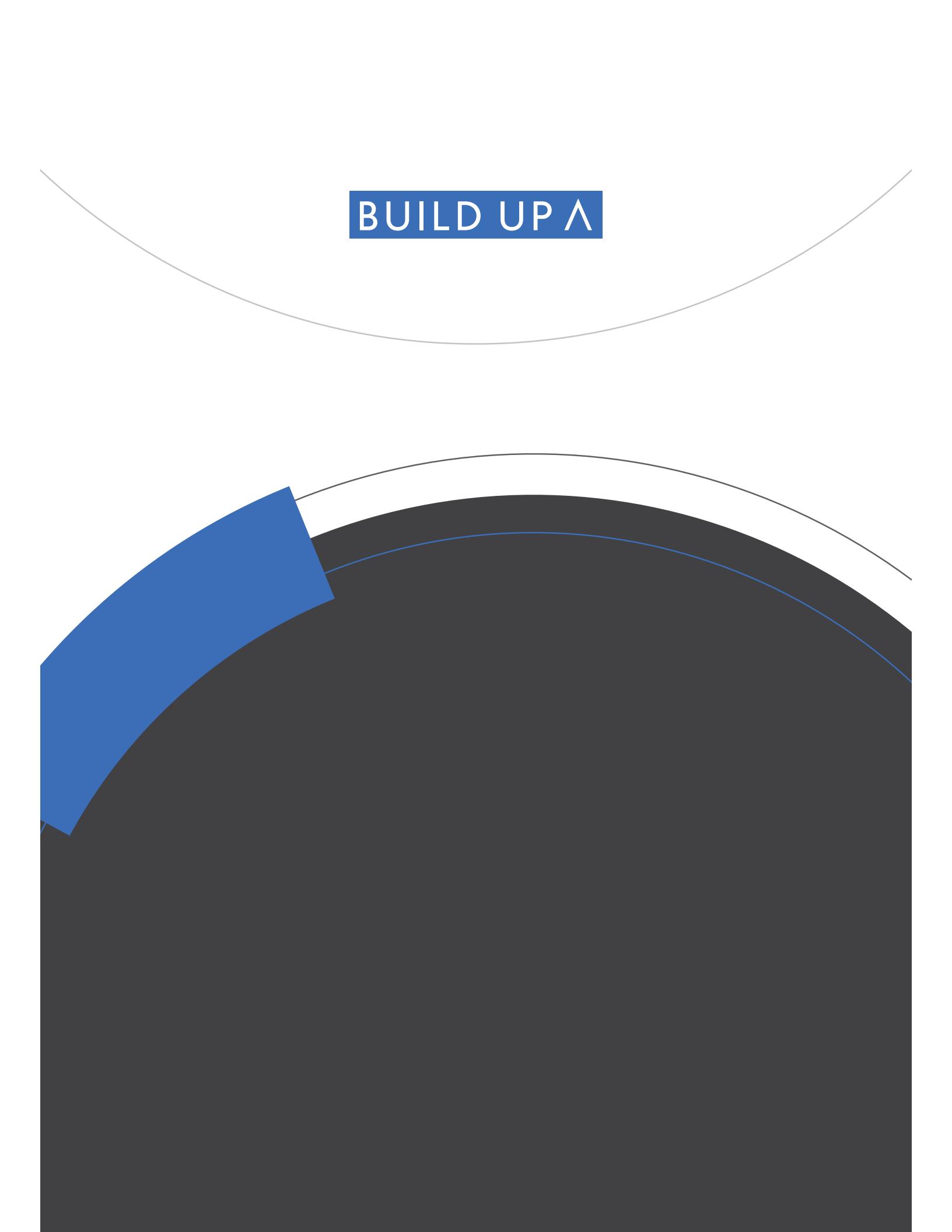
References

- Africa Check. Our Media Literacy Work. Africa Check, 2022. <https://africacheck.org/what-we-do/media-literacy>
- Association of Media Women in Kenya. Laws Governing Media Practice in Kenya: A Journalists' Handbook. AMWIK, 2014. <https://amwik.org/wp-content/uploads/2014/12/Laws-Governing-Media-in-Kenya.pdf>
- Apondu, Charles. Interview. Conducted by Caleb Gichuhi, February 2022.
- Article 19. Kenya Use of Hate Speech Laws. Article 19, 2020.
- Asare, Janice Gassam. Social Media Continues to Amplify White Supremacy And Suppress Anti-Racism. Forbes, 8 January 2021.
- Bloggers Association of Kenya. Petition against the Computer Misuse and Cybercrime Act. BAKE, 2018.
- CFL Advocates. Legal Update: High Court Upholds Constitutionality of Computer Misuse and Cybercrimes Act, 2018. Lexology, 2020.
- Benesch, Susan. Final Report: Countering Dangerous Speech to Prevent Mass Violence During Kenya's 2013 Elections. Dangerous Speech Project, 2014.
- Beykpour and Adegbite. Establishing Twitter's presence in Africa. Twitter, 12 April 2021. https://blog.twitter.com/en_us/topics/company/2021/establishing-twitter-s-presence-in-africa
- Build Up. Case study of #RutoReturnOurFood.
- Constitution of Kenya, 2010. Kenya Law, Accessed February 2022.
- Computer Misuse and Cybercrimes Act No. 5, 2018. Kenya Law, Accessed February 2022.
- Communications Authority of Kenya. What We Do. Communications Authority of Kenya, 2022.
- Communications Authority of Kenya. Guiding Manual for the Collection of Supply-Side Data on Telecommunications/ICT 2019. Communications Authority of Kenya, Accessed February 2022.
<https://ca.go.ke/wp-content/uploads/2020/03/ICT-Indicators-Manual-2019-.pdf>
- Damary, Rita. Kalonzo Wants Uhuru to Drop Pangani 6 Cases to Show True Forgiveness. The Star, 6 May 2018.
- Dangerous Speech Project. What is Dangerous Speech? Dangerous Speech Project, 2022.
- Dangerous Speech: A Practical Guide. Dangerous Speech Project, 2021.
- Dwoskin, Newmyer, and Mahtani. The Case Against Mark Zuckerberg: Insiders Say Facebook's CEO Chose Growth Over Safety. Washington Post, 25 October, 2021. <https://www.washingtonpost.com/technology/2021/10/25/mark-zuckerberg-facebook-whistleblower/>
- Elections Act of Kenya, 2011. Independent Electoral and Boundaries Commission. Accessed February 2022.
- Gadde Vijaya & Beykpour Kayvon. Setting the Record Straight on Shadow Banning. Twitter, 2018.
- Geeska Afrika Online. Kenya to Monitor Social Media During Elections. Geeska Afrika Online, 2017.
- Gichuhi, Caleb. Interview. Conducted by Kate Mytty, February 2022.
- Goldstein and Rotich. Digitally networked technology in Kenya's 2007-2008 post-election crisis. Berkman Center for Internet & Society at Harvard University, 2008. https://cyber.harvard.edu/sites/cyber.harvard.edu/files/Goldstein&Rotich_Digitally_Networked_Technology_Kenya_Crisis.pdf
- International Criminal Court. Rome Statute of the International Criminal Court. International Criminal Court, 1998.

- International Foundation for Electoral Systems. Kenya Electoral Assistance Program. IFES, 2018. <https://www.ifes.org/KEAP>
- Jacobs, Josh. Does Online Hate Drive Anti-Migrant Violence? Financial Times. 2022
- Kachwanya, Kennedy. Interview. Conducted by Caleb Gichuhi, February 2022.
- Karimi, Faith. Twitter Now Speaks Swahili. Poa Sana. CNN, 2018.
- Khan, Irene. Access to Reliable Information Sources Is the Obvious Antidote to Disinformation, Special Rapporteur On Freedom of Opinion and Expression Tells Human Rights Council. United Nations Human Rights, 2021.
- Kemp, Simon. Digital in Kenya: All the Statistics You Need in 2021. Data Reportal, 15 February 2022.
- Kenya Targets Social Media “Bullies” in New Crackdown. Capital News, 4 November 2021.
- Kenya Communications Act, 2008. Kenya Law, Accessed February 2022.
- Kenya Information and Communication (Amendment), Act 2013. Kenya Law, Accessed February 2022.
- Kiss 100 Kenya. Obinna’s Instagram Account Being Shadow Banned. YouTube, 2022.
- Konrad-Adenauer- Stiftung. Kriegler and Waki reports, Summarised Version, Revised Edition. 2009.
- Languages of Kenya. Wikipedia. Accessed February 2022.
- Madung, Odanga, with Brian Obilo. Inside the shadowy world of disinformation for hire in Kenya. Mozilla Foundation, 2021.
- Madung, Odanga. From Dance App to Political Mercenary: How disinformation on TikTok gaslights political tensions in Kenya. Mozilla Foundation, 2022.
https://assets.mofoprod.net/network/documents/From_Dance_App_to_Political_Mercenary.pdf
- Madung, Odanga. Exporting Disinformation: How Foreign Groups Peddle Influence in Kenya Through Twitter. Mozilla Foundation, 2021.
- Mahihu, Morara, and Sambuli. Umati Monitoring Report: July - September 2013. iHub Research, 2013. https://stsinfrastructures.org/sites/default/files/artifacts/media/pdf/jb_umatimonitoringreportjulytoseppdf2013-11-22-13-53-49.pdf
- Media Council Act, 2013. Media Council of Kenya. Accessed February 2022.
- Media Council of Kenya. Media Council of Kenya to Train Journalists on Elections Reporting. Media Council of Kenya, 2022.
- Media Council of Kenya. Complaint no. 001 of 2011, Ambassador Francis Muthaura-vs-the Standard Group and two others, 2011, Accessed March 2022.
- Meta Ad Library Report – Kenya, Political or Issue Ads. Meta, 2022. <https://www.facebook.com/ads/library/report/?source=archive-landing-page&country=KE>
- Meta. Where We Have Fact Checking. Meta Journalism Project, 2022. <https://www.facebook.com/journalismproject/programs/third-party-fact-checking/partner-map>
- Meta. How Facebook’s Third-Party Fact-Checking Program Works. Meta Blog, 1 June 2021. <https://www.facebook.com/formedia/blog/third-party-fact-checking-how-it-works>
- Meta Transparency Center: Government Requests for User Data: Kenya. January – December 2021. Meta, Accessed February 2022. <https://transparency.fb.com/data/government-data-requests/country/KE/>
- Meta. Transparency Center: How Technology Detects Violations. Meta, 19 January 2022. <https://transparency.fb.com/es-la/enforcement/detecting-violations/technology-detects-violations/>

- Meta. How Meta Prioritizes Content for Review. Meta, 26 January 2022. <https://transparency.fb.com/es-la/policies/improving/prioritizing-content-review/>
- Meta. Transparency Center: Taking Action. Meta, 2022. Accessed February 2022. <https://transparency.fb.com/es-la/enforcement/taking-action/>
- Meta. Transparency Center: Hate Speech. Meta, 2022. Accessed February 2022. <https://transparency.fb.com/es-la/policies/community-standards/hate-speech/>
- Meta. Transparency Center: Content Restrictions Based on Local Law: Kenya. Meta, January – July 2021 and July – December 2021. Accessed February 2022.
<https://transparency.fb.com/data/content-restrictions/country/KE/>
- Meta. Transparency Center: Information Requests: Kenya. January – June 2021 and July 2021 – December 2021. <https://transparency.twitter.com/en/reports/information-requests.html#2021-jan-jun>
- Meta AI. ML Applications: AI Advances to Better Detect Hate Speech. Meta, 12 May 2020. <https://ai.facebook.com/blog/ai-advances-to-better-detect-hate-speech/>
- Meta Business. Ads Help Center: Hateful Content. Meta, 2022. <https://business.twitter.com/en/help/ads-policies/ads-content-policies/hate-content.html>
- Meta. Transparency Center: KoPI-G Transparency Report. Meta, October 2021. <https://transparency.fb.com/sr/kopi-report-english-oct-21>
- Metet, Olive (NCIC). Interview. Conducted by Caleb Gichuhi, February 2022.
- Muraya, Joseph. MPs Moses Kuria and Waititu Acquitted of Hate Speech Charges. Capital News, 20 February 2017.
- National Cohesion and Integration Commission of Kenya. Conflict Hotspot Mapping for Kenya. 2022.
- National Cohesion and Integration Commission of Kenya. Police Training Manual: On the Enforcement of the Law on Hate Speech, 2011. National Cohesion and Integration Commission of Kenya, Accessed February 2022.
- National Cohesion and Integration Act of Kenya, 2008. Kenya Law, Accessed February 2022.
- National Cohesion and Integration Commission of Kenya. NCIC Achievements. National Cohesion and Integration Commission of Kenya, Accessed February 2022.
- Nderitu, Alice Wairimu. Kenya: Bridging Ethnic Divides: a Commissioner's Experience on Cohesion and Integration. Mdashalo Bridging Divides Limited, 2018.
- Ogenga, Frederick. Interview. Conducted by Kate Mytty, February 2022.
- Otunga, John and Raashi Saxena. Interview Conducted by Kate Mytty, February 2022.
- Owino, Sheila Akinyi. Interview. Conducted by Kate Mytty, February 2022.
- PeaceTech Lab and Mercy Corps. Election Violence Prevention. PeaceTech Lab, 2017. <https://www.peacetechlab.org/prevent-election-violence>
- Penal Code of Kenya, Revised 2012. Kenya Law, Accessed February 2022.
- Perrigo, Billy. Inside Facebook's African Sweatshop. Time, 18 February 2022. <https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/>
- The Sentinel Project. The 2022 Kenyan Election. The Sentinel Project, 2022. <https://thesentinelproject.org/project/kenya-election-2022/>
- Shiundu, Alphonse (Africa Check). Interview. Conducted by Caleb Gichuhi. March 2022.

- Sochin. Social Media Influencers, Hate Speech Monitoring, Early Warning & Response. Sochin, 2018. <https://www.sochin.agency/our-work>
- State of Internet Freedoms in Kenya. Collaboration on International ICT Policy in East and Southern Africa, 2014.
- Súilleabáin, Andrea Ó (editor). Leveraging Local Knowledge for Peacebuilding and Statebuilding in Africa. New York: International Peace Institute, March 2015. https://www.ipinst.org/wp-content/uploads/2015/03/IPI-E-pub-Leveraging_Local_Knowledge.pdf
- Twitter. Help Center: Abusive Behavior. Twitter, 2022. <https://help.twitter.com/en/rules-and-policies/abusive-behavior>
- Twitter. Help Center: Hateful Conduct Policy. Twitter, 2022. <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>
- Twitter. Help Center: COVID-19 Misleading Information Policy. Twitter, 2022. <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>
- Twitter. Help Center: Civic Integrity Policy. Twitter, 2022. <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>
- Twitter. Help Center: Our Range of Enforcement Options. Twitter, 2022. <https://help.twitter.com/en/rules-and-policies/enforcement-options>
- Twitter. @TwitterMoments. Accessed February 2022. https://twitter.com/TwitterMoments?ref_src=twsr%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor
- Twitter. Information Requests: Kenya, January – June 2021. Twitter. <https://transparency.twitter.com/en/reports/information-requests.html#2021-jan-jun>
- United Nations, 2022. Understanding Hate Speech, Accessed February 2022.
- United Nations. Convention on the Prevention and Punishment of the Crime of Genocide. United Nations, 1948.
- United Nations. International Covenant on Civil and Political Rights. United Nations, 1966.
- United Nations, Office of The High Commissioner for Human Rights. International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). United Nations, 1965.
- Wanjohi, Grace. Interview. Conducted by Kate Mytty. February 2022.
- Williams, Matthew L., Pete Burnap, Amir Javed, Han Liu, and Sefa Ozalp. "Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime." *The British Journal of Criminology* 60, no. 1 (2020): 93-117.
- Zuckerman, Ethan. Kenya: Citizen media in a time of crisis. Ethan Zuckerman Blog, 20 June 2008. <https://ethanzuckerman.com/2008/06/20/kenya-citizen-media-in-a-time-of-crisis/>



The background features a large, dark gray circle with a thin blue outline. A thick blue wedge is positioned on the left side of the circle, extending from the bottom-left towards the top-left. Above the circle, two thin, light gray curved lines form a downward-opening parabola shape.

BUILD UP A